OBITUARY NOTICE

PROFESSOR SAMUEL ALEXANDER, 1859–1938

Professor Samuel Alexander, who had held the Sir Samuel Hall Chair of Philosophy at Manchester University from 1893 to 1924, died at his home at Withington, Manchester, on 13 September 1938, at the age of seventy-nine.

He was born in Sydney, studied at Wesley College, Melbourne, and Melbourne University, and came in 1878 to Balliol College, Oxford. Here his thinking was influenced, as his earlier work shows, by Green and Nettleship. After graduating he was made a Fellow of Lincoln College, where he remained as philosophy tutor from 1881 to 1893. Here he developed his interest in the empirical aspect of psychology, and enriched Oxford's remarkable collection of legends about this study by a story concerning a queer fellow who strapped people to machines intended to measure emotions, and once bounded into the experimental room with the kindest of faces, covered with ornamentations to induce terror. This, however, was not the emotion produced in the victim.

During these years, under the influence of Green, he wrote *Moral Order and Progress*, which, a critic writes, already showed characteristics of Alexander's mature mind; a great respect for fact, and a love of synthesis. In 1893 he succeeded Adamson at Manchester, which was then Owens College. The evolution of Manchester University owes much to Alexander's unique fusion of simplicity, modesty, good fellowship and business ability. Those were the days of trams and bicycles, when the nearby suburbs housed the University staff. The internal combustion engine had not yet begun its calamitous destruction of culture, of which the centrifugal trek from the cities was an early feature. In such a *milieu* of easy "dropping-in", Alexander's influence was all-pervasive. His imperfect hearing, even at that time, was a handicap, but not entirely, for people delighted to speak directly to him. He ignored idle chatter, and sometimes seemed psychically deaf even to relevant questions, proceeding conversationally with the unhurried dignity of a liner entering dock. Perhaps he was aware of this advantage and, too, of the mental mechanism of projection, for once he opened a public discussion by remarking: "When I said this in London, I think Mr — (a dissident philosopher) did not quite hear some of my objections to his views. To

ensure that he shall hear properly to-night, I present him with a copy of my paper."

At first, in Manchester, Alexander published little. Writing was impossible when he was sloughing the skin of idealism which had grown upon him in Oxford. Idealists had argued that to be known is to be dependent on mind. In 1905, Alexander began to counter this with assertions which seemed to be sheer paradox, for the Gifford Lectures and their fuller statement, *Space, Time and Deity*, were still below the horizon. To be known, said he, was to be independent of mind. In exemplifying this assertion, the stately trees in Brunswick Road (for this reason alone they ought to be preserved) must have been utilized almost daily.

Alexander's way of exposition was general and abstract. It consisted of cheques drawn upon the bank of knowledge with a masterly, confident, if illegible hand. He vouchsafed the reader few concrete examples, and these were casually lifted from the conventional philosopher's portable pack of trees, tables and teacups.

Perhaps this way of writing led Alexander farther away from the average reader than either of them ever suspected. It is difficult to believe that he noticed the crudity of early behaviourism. In the exciting years, for psychology, of 1910–14, he seemed almost bored by discussions of Freud. "The stuff that dreams are made on", was, in his view, the stuff that physicists ought to study, though he knew them too well to believe that they ever would. He seemed unmoved by speculations about the mechanism of dreams. Two optical images can be superposed by mirrors; what then is mysterious about Freud's 'Condensation'? Alexander even denied the uniqueness of the relation of knowing. He once said, "The teacup knows the table, but it does not enjoy itself; I know the teacup but I also enjoy myself." That was all the difference, for ordinary metaphysicians, but for the sociable Alexander, what an all!

The contrast between his teaching and that of Külpe at Würzburg or Freud in Vienna would have been a shock to any Continental scholar. Alexander, perhaps urged by his Australian blood, was establishing a lonely frontier post of psychology, and looking away from it across a desolate No-man's Land to the adjacent territories of physics, metaphysics and aesthetics. In logic and ethics he now took almost a perfunctory interest.

To understand him between 1910 and 1914, one would have to comprehend Alexander's Manchester, which, to borrow an American description of another place, was not so much a city as a state of mind. That state of mind was almost dominated by three influences from the Anti-

podes: Alexander, Rutherford, and Elliot Smith. They had slipped into Manchester's culture pattern as easily and naturally as into its University common room.

Manchester's straight talk, her ferocious contempt for appearances, her unconcealed uninhibited friendliness for people she liked, her gentle cherishing of certain cultural values—having made sure that a few extraverted business men were willing and able to pay for them—her local sport of bubble pricking (or bashing), her proud, even defiant cosmopolitanism; with all these notes this trio could harmonize. Alexander's spectacular appearance and cathedral-organ voice, Rutherford's bucolic bellow, Elliot Smith's illusory shyness: without them Manchester's university life cannot be pictured.

It is as difficult to estimate Alexander's influence upon psychology as it is to limit an account of his influence upon psychologists. The central point of his teaching was that a completely conational psychology was possible, that the alleged cognitive element is non-mental, and that feeling is a qualification of conation. Perception can be resolved into act and external thing. This is not the result of common-sense, nor is it naïve realism. That it happens is not a postulate but a description of the event in its simplest terms. On the different levels of experience, perception, imagination, memory, etc., there corresponds to each form of conation a certain form of non-mental object or cognition. Both the object cognized, and even the organic and kinaesthetic *sensa* in which attention results, are objects contemplated. They are not enjoyed, undergone, lived through, or suffered, as is the mental act.

This 'act-object' contrast pervades all his teaching. It is the business of psychology to describe and to analyse enjoyments; this is introspection. The subject, as enjoyed, is the continuous tissue of its acts of conation; consciousness is the general form of such enjoyment. The combination of the subject, who thinks, feels or wills, and the bodily self, he called the person, and distinguished stages in the growth of personality.

Alexander's special interests in acts rather than in objects (in '-ings' rather than in '-eds', to quote Prof. C. Lloyd Morgan) perhaps made him uninterested in those facts which follow the behaviourist around, like neglected animals, with sad, reproachful eyes; synaesthesia and eidetic images, memory images, hallucinations, dreams; indeed, one of Alexander's greatest admirers wrote of his "mistreatment of imagery in memory". Remembering was for him an act of desire directed backwards towards the past. The image, if the original object had been a physical one, is physical also; is the same thing as the percept, appearing in a

different form. Deciding that half the problems of the psychologist concerned non-mental happenings, and having handed over the mandate to the unheeding physicists, he busied himself with other matters, and consequently his references to problems of memory seem a little weary and elusive. For the last thirty years of his life Alexander's interests were chiefly in metaphysics, though he wrote actively upon aesthetics, about which the present writer is utterly incompetent to judge. But Alexander was the kindest of critics, and so often used the term 'beautiful', especially of human beings, that his essays on beauty were also essays in benevolence.

If Alexander's character was formed in the southern hemisphere, Oxford moulded his personality, bestowing that combination of sonorous public speaking, impressive diction, and occasional insouciance about its comprehensibility.

Alexander was a mass of contradictions, resolved like the discords in a symphony. Careless of externals, yet amiably consenting to be photographed, drawn, painted, etched and finally sculptured, garbed in a manner which launched a thousand legends, but loving his finery as Public Orator; magnificently declaiming his speeches, composed with the greatest artistry and after patient researches into the graduand's public and private records, yet capable of thanking a beloved friend publicly in a lazy tangle of *clichés* and repeated half-sentences, with naughty beams of affectionate atonement; a lover of all nations, but not a traveller, looking like a cartoonist's dream of a professor, yet punctual and punctilious, unforgettably inspiring, yet producing only a few distinguished students, loving and beloved by man, woman, child and animal, yet lonely, and 'rather liking it'.

T. H. Pear.

Prof. WILLIAM McDOUGALL, 1871–1938

OBITUARY NOTICE

PROFESSOR WILLIAM McDOUGALL, 1871–1938

With the death of William McDougall on 28 November 1938, British psychology lost one of its foremost representatives and the British Psychological Society one of its most distinguished Honorary Members. He was an original member of the Society, the first meeting of which (on 24 October 1901) was, indeed, preceded by a number of informal discussions of a little group which met in his laboratory at University College, London. He acted as Hon. Treasurer to the Society from 1905 to 1908 and again from 1911 to 1916, making during all this time frequent contributions to the Society's meetings. He was an Associate Editor of this *Journal* from its inception, and in 1934 was elected an Honorary Member of the Society, when in that year it was decided—for the first time since the War—that the short list of those whose services to psychology the Society wished to recognize in this way should include representatives of British as well as of foreign nationality.

Born in Lancashire on 22 June 1871, McDougall early showed himself to be the possessor of exceptional abilities, and, as he tells us (in his illuminating contribution to *Psychology in Autobiography*), it was soon made plain to him that he was expected to distinguish himself. After attending a private school, he spent a year at the *Realgymnasium* at Weimar and then, when only fifteen, entered the University of Manchester, 'a serious student', but one whose physical endowments equalled his mental capacities, and who found time to represent his University at Rugby football, on the track and on the river. Four years at Manchester were followed by four at St John's College, Cambridge, and these in turn by yet another four at St Thomas's Hospital, London, towards the end of which period he published his first papers, which were devoted to the subject of muscular contraction. He had no intention, however, of practising medicine, and even as an undergraduate had determined to devote himself to a study of the brain and nervous system, "in which were locked the secrets of human nature", though a reading of James's *Principles of Psychology* during his hospital years revealed to him another and complementary method of approach. After election to a fellowship at his own College at Cambridge, he joined the Cambridge Anthropo-

logical Expedition to Torres Straits, then proceeded to Borneo at the invitation of Dr Charles Hose (in collaboration with whom he subsequently produced a large two-volume work on *The Pagan Tribes of Borneo*).

Returning to Europe in 1900, he married Anne Hickmore of Brighton and spent some time at Göttingen with G. E. Müller, where he embarked on the experimental study of vision that occupied much of his time during the next few years. In 1900 he was invited by James Sully to undertake the teaching of experimental psychology at University College, London, where a small laboratory had recently been established (equipped with the apparatus formerly used by Münsterberg at Freiburg), and he continued to teach there till 1906, though in 1904 he entered upon the more important duties of Wilde Reader in Mental Philosophy at Oxford. He was debarred by the conditions of his Readership from dealing in his official lectures with the experimental approach, but Francis Gotch, the Professor of Physiology, put at his disposal, first one, and then (when a new wing was added to the physiological laboratory) six rooms, where, in a private capacity, he was able to continue research, and where he gathered about him a small group of students, including M. E. Bickersteth, William Brown, Cyril Burt, Horace English, A. M. Hocart, May Smith and the present writer. To the members of this group his teaching and companionship were equally delightful. He himself tells us that he regarded them as "brands plucked from the burning", and these words indicate a certain inevitable dissatisfaction with the conditions at Oxford, where psychology had no place in the curriculum, where potential students were discouraged by philosophers in whose opinion psychology could not and therefore did not exist, and where the scientists on their part (with a few exceptions) also regarded with grave suspicion the upstart branch of knowledge which he represented. There were, however, certain compensating advantages, and he made full use of the freedom he enjoyed to lecture on any aspect of psychology he chose, much of his later published work being adumbrated in these early lectures. The informal meetings in the laboratory were, however, for those who were privileged to attend them, a more potent and inspiring influence than the official lectures, although these latter sometimes drew good audiences (especially a course of hypnotism, which was regarded as a very daring subject for a University lecturer to tackle).

McDougall's earlier publications clearly reveal his interest in both the neurological and the psychological approaches to the problems of the mind. Indeed his very first contributions of a psychological nature (three articles in *Mind* of 1898) were directly concerned with problems of psycho-

physical correlation. There followed a brilliant series of papers, many of them on vision, others on the 'attention process', hypnotism, fatigue, and other subjects, all largely 'psychophysical', in the sense that introspective observations (often based on ingenious experiments) were interpreted to a great extent in physiological terms. His first book, the small *Physiological Psychology* published in 1905, proved a very fascinating general exposition from this point of view and (as the present writer can testify) exercised a great influence on many students in Oxford and elsewhere who were otherwise suspicious of psychology. From this time onwards, however, his interest in neurology and his satisfaction with physiological explanations appear to have waned, giving place to a more biological point of view—his biology, from the start, being dominated by the concept of purposive adaptation. In this change of standpoint we may perhaps trace the conflicting influences of Sherrington and James respectively, both of whom were of great importance in his life, though the inspiration derived from James was destined to have the more permanent effect. Anyhow, from the time of writing his *Body and Mind: A History and Defense of Animism* (1911), McDougall was but little occupied with problems of brain and nerve function. And this book itself stated boldly and clearly the thesis that mind itself must be regarded as a potent cause in development and evolution. By general admission it presents by far the strongest case that has ever been made for Interactionism, and it seems likely to remain a work of supreme importance in this field until further facts of fundamental significance as regards the mind-body relationship shall have been discovered.

Meanwhile, with his second book (the *Introduction to Social Psychology*, 1908) McDougall had entered a new and more purely psychological field, and had definitely emerged as champion of what subsequently became known as the hormic theory, the view that stresses the striving and purposive aspects of all life and that seeks to trace the energy displayed in every activity of living creatures to some inborn disposition or instinct. The book will (as he himself recognized) almost certainly be reckoned his most original contribution to psychology. At any rate, it soon became, and continues even now after more than thirty years, to be one of the most popular books on psychology. Not only did it complete the task (begun by Ward and Stout) of correcting the intellectualist and hedonist bias of the associationist views then dominant both inside and outside the laboratory, but at one stroke it produced an orderly scheme for the understanding of orectic life from its simplest to its most complex manifestations, a scheme which he himself was able to use in all his

subsequent writings and which has also been adopted in whole or in part by many other authors.

In the title of this work he had used the term *Introduction* in the rather unusual sense of Prolegomena, and he regarded his study of instinct, emotion, habit, sentiment, etc., as the essential preliminary to a treatment of social psychology itself, which was to constitute his *magnum opus*. The war, however, interrupted his plans, and after the publication of one other small book, *Psychology, the Study of Behaviour* (which also proved extremely popular), he found himself in 1914 driving an ambulance on the Western front, and then—more appropriately—a major in the R.A.M.C., dealing with cases of 'shell-shock'. This work continued till 1919 when, feeling that he needed to digest his material on the functional nervous disorders (which now filled thirty-five notebooks) in the light of the experience of others, he visited C. G. Jung at Zürich and was for a time analysed by him. He came away "enlightened but not convinced", and his book on *Abnormal Psychology* did not appear till 1926, though in the interval he wrote two other works of great importance —*The Group Mind* (1920) and the *Outline of Psychology* (1923). The former was intended to be the first volume of the great projected work on social psychology already mentioned, but he was disappointed at its reception and "the *magnum opus* went a-glimmering", though he subsequently returned to social problems in several minor works; the second was in effect the working up of the principles enunciated in the *Introduction to Social Psychology* into a general system of psychology, into a scheme, as he believed (and as many will agree), "more complete than any other".

Before this, however, McDougall had moved from England to the United States, having in 1920 accepted an invitation from Harvard to occupy the Chair of Psychology left vacant by the death of Münsterberg. As regards prestige, equipment, and number of students, Harvard offered vastly greater opportunities than Oxford; but soon after his arrival McDougall found himself in a behaviouristic atmosphere that was distinctly uncongenial. It is true that he himself had called psychology 'the study of behaviour', but to him behaviour meant conduct determined by purpose and desire, whereas to many of his colleagues and students it meant a materialistic view of life conceived in terms of stimulus-response mechanisms and conditioned reflexes. For this and other reasons, McDougall never fitted quite happily into the American scene. He tells us, however, that he never regretted the adventure, and indeed, as he himself recognized, there was an element in his character which

made it impossible for him to feel entirely at home in any group or any environment. Though he was very far from being in all respects a radical, he was endowed with a certain intellectual 'arrogance' combined with a tendency to scepticism, as a result of which he nearly always found himself in opposition to some prevailing tendency or theory. In the early days he combated the fashionable theories of Hering in the field of vision and the even more fashionable solution of the mind-body problem in terms of psycho-physical parallelism, and it was perhaps a tragedy that in one of the few important matters in which he thought he could wholeheartedly adopt the views of another worker (in the case of Shand's doctrine of sentiments), he soon found that there existed irreconcilable differences of opinion which he had not at first suspected. From early years onwards he became increasingly sceptical of the prevailing Weismannism, which he found 'all-dominant' even in his Cambridge days and the influence of which he suspected of being due to its association with 'mechanistic dogma in biology'. In America he resolved to put the opposing Lamarckian theory to the test and began the long series of experiments with white rats that is still in progress—experiments designed to show whether, as Lamarck had supposed and as Weismann had denied, there really existed any tendency for 'acquired characters' to be inherited. The experiment, as is now well known, has produced results pointing clearly in the Lamarckian direction, though their precise significance is still a matter of dispute. Whatever the ultimate outcome, however, the experiment was devised and carried out on a heroic scale and it may indeed yet come to be regarded as of epoch-making significance.

His tendency to oppose prevailing views or theories continued unabated into later life, and indeed many of his last writings were polemical in nature. Behaviourism, psycho-analysis, Gestalt, each in turn became an object of detailed and penetrating criticism. His battle with behaviourism was the longest and bitterest of all, partly because American psychology was so largely influenced by this doctrine in the years following the war, and partly because behaviourism seemed to him to be utterly opposed to his own conceptions of life and mind and based on assumptions which threatened to deprive psychology of all its true significance. In both Gestalt and psycho-analysis there was much that he admired, in particular the new approach to old problems, the ingenious experimentation and the rejection of hedonism and associationism (in the case of Gestalt) and the whole-hearted adoption of the categories of wish or purpose (in the case of psycho-analysis). Moreover McDougall was impressed by the successful application of psycho-analytic concepts to social

phenomena and regarded this as one more proof of the correctness and fruitfulness of the hormic view in psychology. But he deplored the vague and, as it seemed to him, utterly inadequate physical conceptions of energy prevalent among Gestalt psychologists; while the psycho-analysts on their part offended him with what appeared to be their inadequate battery of instincts, their overworking of the sex instinct, their hypostatizations, their uncritical resort to psychological hedonism and their somewhat chaotic theorising generally. It is true that much of psycho-analytic theory does appear to be a clumsy makeshift when compared with McDougall's own neat and systematic formulations, though at the same time few will deny that as regards the detailed understanding of individual motives McDougall did not display so deep a penetration as did Freud. Perhaps, in his heroic endeavour to bring order out of chaos, the system he devised was *too* neat and precise to serve as an adequate theoretical framework for the vast complexities of human motivation and behaviour. Indeed McDougall himself has, during the last fifteen years, introduced various extensions and modifications of his original system (e.g. his doctrines of 'temper', 'taste' and the 'derived emotions'), which have given it a less orderly appearance. Nevertheless, some kind of consistent theory was, and still is, an urgent necessity for the fuller understanding of orexis, and McDougall's formulations, whatever further revision they may ultimately require, have beyond all doubt been of immense service to contemporary psychology.

In 1927 McDougall moved from Harvard to Duke University. In his new position he was able to spend the early summer months in England. But whether he was officially in term or in vacation, he was always busily at work, and during the last years books and articles continued to appear, many of them minutely critical, as we have already indicated, others broadly constructive, in the sense of developing the implications and applications of his own purposive view of psychology. Among works of the latter kind were *Character and the Conduct of Life* (1927), *Modern Materialism and Emergent Evolution* (1929), *World Chaos* (1931), *Religion and the Sciences of Life* (1934), *The Frontiers of Psychology* (1934), while *The Energies of Men* (1932) was a restatement of his own position from the point of view of pure psychology. His last work, *The Riddle of Life*, has only just appeared, having been completed after his partial recovery from a severe operation in February 1938.

Apart from what may be regarded as his main activities, McDougall had a number of sidelines in which he also displayed his characteristic energy. He was a keen eugenist and he developed this theme in *National*

Welfare and National Decay (1921) and in several other writings. In *Ethics and Some Modern World Problems* (1924) and in *Janus or the Conquest of War* (1927) he made the interesting suggestion that the national ownership of aircraft (or at any rate fast flying aircraft) should be forbidden, their use being confined to the purposes of an international police force. In view of subsequent events we may perhaps regret that the world did not see its way to treat this proposal with greater seriousness. He also throughout his life displayed an open-minded interest in psychical research, feeling that it was a scandal that psychologists should refuse to lend a hand to the very small body of workers who were really seeking the truth in this much debated field. In this connexion he sponsored the experiments of his colleague, Prof. J. B. Rhine, on Extrasensory Perception, which have recently aroused such widespread attention.

McDougall was a member of very many scientific bodies and societies, both in Europe and America. Before all of them he read papers at one time or another and in many cases he served in some capacity on their governing bodies. He became a Fellow of the Royal Society in 1912, a Fellow of St John's College, Cambridge in 1898, and of Corpus Christi College, Oxford in 1912, while in 1919 he received the Honorary D.Sc. of Manchester, his own first University.

In spite of his great ability, his productiveness and energy, and his undoubted achievements, in spite even of what he himself has called his 'arrogance', McDougall was not free from disappointments and from frequently recurring doubts as to the value of his own work. Compelled, probably by his nature, to play for the most part the role of a lone fighter, he founded no school, as so many other modern psychologists of comparable eminence have done, though this fact is probably also due in part to an adverse intellectual atmosphere (at Oxford and at Harvard) and to the circumstance that his psychology was not especially associated with any particular method or technique of investigation. He sometimes felt very keenly the absence of any recognized band of followers who accepted the main principles he had laid down. He was even at times (as some thought) unduly sensitive of what he considered to be the neglect of his teaching on the part of colleagues and students and their failure to make adequate use of it in their own contributions. He perhaps failed to realize the full degree of the influence he was exerting, overestimated the differences between himself and others, and saw stark incompatibilities in certain cases where further reflexion might have revealed possibilities of bridging the gap between his views and those of fellow-workers. But

such doubts and misconceptions, in so far as they existed, were rarely if ever allowed to interfere, either with his personal relationships, which, if seldom intimate, were almost invariably friendly, or with steady, persistent work at the tasks which he had set himself. The disappointments, the difficulties, the doubts, such as they were, were for himself alone; his achievements were for the benefit of his chosen science, and through that for mankind at large. For he was acutely conscious of the important role that psychology ought to play in human affairs, and at the same time regretfully aware that it was at present unfitted to play this role, partly because of its humble and backward position in the hierarchy of sciences, and partly because of dissensions among psychologists themselves. McDougall's life exemplified his teaching: it was a life of single-minded purpose, which guided him throughout the innumerable details of each individual undertaking. Early in his career he determined to devote himself to pure research into the secrets of human nature: this course he consistently pursued until the end. Few men could say with the same full sincerity as he: "In the main I have lived the sort of life which in my youth I judged to be the most desirable: and that perhaps is all a man can properly demand."[1]

J. C. Flugel

[1] In preparing the above account, I have enjoyed the great advantage of consultation with Prof. Cyril Burt and Dr May Smith, to both of whom I am much indebted for kind help and comment.

THE CO-OPERATION OF SOCIAL GROUPS
PART II

By D. M. CARMICHAEL

From the Cambridge Psychological Laboratory

IV. *Further experiments in constructive social thinking* (pp. 329–331).
(i) *Material of Case II:* "*Co-operation in agriculture*" (pp. 330–331).
(ii) *Method of presentation* (p. 331).
V. *Analysis of results of Case II* (pp. 332–334).
(1) *Numerical analysis* (p. 332).
(2) *Qualitative analysis* (pp. 333–334).
(i) *Permanent co-operation* (pp. 333–334).
(ii) *Breakdown* (p. 334).
VI. *Case III* (pp. 335–336).
(i) *Material of Case III:* "*An anachronism removed*" (pp. 335–336).
(ii) *Method of presentation* (p. 336).
VII. *Analysis of results of Case III* (pp. 336–339).
(1) *Numerical analysis* (pp. 336–337).
(2) *Qualitative analysis* (pp. 337–339).
(i) *Increasing co-operation* (pp. 337–338).
(ii) *No real co-operation* (p. 338).
(iii) '*Doubtfuls*' (pp. 338–339).
VIII. *Summary of results of Cases I, II and III* (pp. 339–340).
IX. *Discussion* (pp. 340–344).
(1) *Material* (pp. 340–342).
(i) *The questionnaire* (pp. 340–341).
(ii) *Detailed and generalized descriptions* (pp. 341–342).
(2) *Some possible criticisms* (pp. 342–344).
(i) *Concreteness* (pp. 342–343).
(ii) *Psychological remoteness* (p. 343).
(iii) *Real and described situations* (pp. 343–344).
X. *Summary* (p. 344).

IV. Further experiments in constructive social thinking

In the case already considered (this *Journal*, xxix, 206–31) conditions which might be held to facilitate or to hinder the co-operation of social groups were presented in a personal and particular form. The two instances which remain to be discussed present evidence for judgement in

a more generalized form. The analysis of the results secured can be set out briefly, and it will be seen that interesting differences are found between these cases and the earlier one.

(i) *Material of Case II: "Co-operation in agriculture"*

Yesterday the culminating step was taken in a series of negotiations which, as most people know, have been steadily progressing for several months. Mr Ardern, representing the Union of Agricultural Engineers, and Mr T. Smith, of the Amalgamated Association of Agricultural Labourers, set their signatures to Articles of Agreement which, it is hoped, will bring these powerful, but widely different groups into close and fruitful co-operation. In many quarters it has for long been maintained that the different organized sections of any complicated industry must be brought to unite if the best interests of each section are to be secured. It is undeniable that in the past the increasing invasion of agriculture by the skilled technical worker has produced a great amount of irritating friction. In the mass, the agricultural labourers of this country represent a solid, but excessively conservative element of the community. The alert engineer, on the other hand, marches in the van of an incalculably rapid and bewilderingly varied progress, and is perhaps apt, on that account, to be of a radical frame of mind which welcomes change, and tends to regard attachment to ancient methods and practices as timid and bovine....

A survey of the Articles of Agreement makes it clear that one of the outstanding hopes of those who have worked for this union of groups is that thereby the bargaining power of the labouring class in agriculture towards securing for themselves better economic conditions will be enormously increased. Indeed, a superficial study of the Articles leaves one with the impression that the Association of Agricultural Labourers have far more to gain from the proposed co-operation than have the Union of Agricultural Engineers. But deeper considerations make it obvious that no such impression can be long maintained. Ultimately the interest of all sections of any organized industry must be congruent, and any amelioration of the conditions of any one section must work for the betterment of all. This fact, however, is frequently obscured, and the achievement of this desirable end is hindered, by sectional jealousy and the biassed demands of short-sighted agitators. It is a welcome sign of the times that the Agricultural Engineers clearly recognize that their own good fortune must be founded upon the bed-rock principles of justice and cannot be finally and firmly secured save by active concern for the lot of their less fortunate brothers in industry....

Already critics are not lacking who endeavour to point out that these Articles contain the seeds of division. They say, with undoubted truth, that the mode of organization of the Engineers differs greatly from that of the Labourers. The former, with their vastly greater internal differentiation, have developed along lines which allow far greater scope and power to the expression of local opinion. Their leaders are democratically elected and are, in a true sense as regards their policy, the servants of those whom they lead. The latter, varying far less among themselves, and from district to district, have never developed local organization to a high degree. The Labourer leaders have come to the front mainly on a basis of their particular personal qualities, have framed their own policy and maintained it, when they have been successful, primarily because they possess the power to persuade or to command.

It is maintained—perhaps we should say rather it is hoped—that co-operation between groups whose organization is on the one hand democratic and on the other despotic can never succeed.

Opponents of the Agreement have even sought the aid of 'science'. They have called in the ubiquitous psychologist. Apparently some eminent and ingenious person, armed with methods which it is not easy to understand, has 'demonstrated' that the 'average I.Q.' level in the Union of Agricultural Engineers is 160, whereas the corresponding measure in the case of the Amalgamated Association of Agricultural Labourers is 90. From these mystic figures dire results are prophesied.

The essential rightness of the Agreement remains. We are convinced that we speak for the solid mass of public opinion in this country when we assert that its principles will stand firm against all selfish and narrow-minded attack, an object-lesson of the way in which industrial organization should and must move, not only in this country, but throughout the world.

This passage was written by Bartlett in the form of a hypothetical leading article, and was presented as such to all subjects. In view of the interesting results which it had already yielded,[1] we decided to test it more extensively.

It differs in three obvious ways from Case I: (1) no specific setting is given; (2) instead of a narrative of events, with little explanation, we have a deliberate and generalized statement of possible leading conditions; (3) the groups described are not directly in contact but attempt to secure co-operation through representatives.

(ii) *Method of presentation*

This material was presented to two groups, both academic—one a mixed class of psychology and anthropology students in Cambridge, and the other a class of psychology students in Edinburgh. The average age distribution in both groups was closely similar.

Bartlett's method of presenting the material was followed, so as to justify comparison of results. The passage was read once to each group as a whole, and any subject who wished to do so was allowed to refer to the text as much as he pleased in giving his answer.

The instructions which followed the reading were these:

Assuming the facts are as reported, and ignoring any *logical* excellencies or failings in the argument, try to base an opinion, based as far as possible on *psychological grounds* only, as to whether the co-operation sought between these two groups is likely to be successfully maintained, or must inevitably break down. State the basis of your opinion in psychological terms, but without argument.

[1] "The co-operation of social groups" (1938), *Occupational Psychol.* XII, 30–42.

V. Analysis of results of Case II

(1) *Numerical analysis*

Table XIV

Group	Permanence	Breakdown	Conditional	Total in group
I. English Academic	15	30	—	45
II. Scottish Academic	40	70	5	115
Total	55	100	5	160

Table XIV gives a simple numerical analysis of the results of Case II, and Table XV the same results expressed as percentages.

Table XV

Group	Permanence %	Breakdown %	Doubtful %
I. English Academic	33·3	66·7	—
II. Scottish Academic	34·8	60·8	4·4
Both groups taken together	34·4	62·3	3·3

The similarity in distribution in these two groups is very striking, and is almost identical with that which Bartlett obtained using the same material. From two English groups, one academic and the other a W.E.A. group (consisting of persons without University training, but with a wider experience of practical affairs), Bartlett got an overall distribution of 35% for permanence, 61% for failure, and 4% doubtful. There was, however, an interesting difference between his two sets of subjects. In the University group opinion was exactly equally divided between probable breakdown and probable permanence. In the practical group the figures were—for permanence 20%; for breakdown 80%. No member of the practical group adopted a neutral attitude. Whether a further extension of the survey so as to include a group corresponding to this practical group of Bartlett's would significantly alter the overall distribution of opinion has still to be determined.

In the Scottish groups there is one small internal difference which seems worth noting: of the 70 students who predicted failure, 40 were men and 30 women; and of the 40 who predicted permanence, 17 were men and 23 women—i.e. the women were more inclined than the men to predict permanence, and less inclined to predict breakdown.

(2) *Qualitative analysis*

(i) *Permanent co-operation.* In Table XVI are classified the reasons given for predicting *permanent* co-operation. Table XVII shows the relative emphasis on each reason calculated as described above (p. 219).

Table XVI

Reason	E.A. (15)	S.A. (40)	Both groups (55)
I. Overlap of economic interests	12	26	38
II. Intelligence of engineers	3	9	12
III. Workers of the world uniting	—	6	6
IV. Education reducing difference in I.Q.	—	4	4
V. Effective contact only between leaders (difference in I.Q. small)	—	3	3
VI. Difference in I.Q. irrelevant	—	8	8
VII. Agreement once reached unlikely to break down	—	3	3
VIII. Other 'isolated' reasons	9	3	12
Total	24	62	86

Table XVII

Reason	E.A. %	S.A. %	Both groups %
I	50	41	44
II	12	15	14
III	—	10	7
IV	—	6	5
V	—	5	4
VI	—	13	9
VII	—	5	4
VIII	38	5	14

As in Bartlett's results (when 55% of the subjects who predicted permanence did so on the ground that an overlap of economic interests would defeat all tendency to disruption), the economic factor is by far the most heavily weighted. The superior intelligence of the engineers is considered likely to put them in a dominant position, but no mention is made, in any of the answers analysed above, of a "swing towards democracy among the labour class", such as Bartlett's subjects predicted. Instead, the engineers are expected to adopt a rather paternal attitude, continuing to concede all advantages to the labourers, and showing patience and adaptability.

In Factor VII we see employed the generalization examined in Case I above.

The 'negative reason' that difference in I.Q. is irrelevant, takes two forms: (*a*) intelligence tests are unreliable, and (*b*) differences in intelligence are not usually an obstacle to co-operation.

The 'isolated' reasons included under VIII are reasons which occur only once and are thus unclassifiable.

(ii) *Breakdown.* In Table XVIII are given the reasons on the basis of which breakdown is predicted, and in Table XIX the relative emphasis on each reason.

Table XVIII

Reason	E.A. (30)	S.A. (70)	Total (100)
I. Difference in intelligence	18	35	53
II. Differences in organization and methods of representation	22	16	38
III. Difference in way of life and general outlook	12	60	72
IV. Conflict of interests	8	26	34
V. Other 'isolated' reasons	—	3	3
Total	60	140	200

Table XIX

Reason	E.A. %	S.A. %	Both groups %
I	30	25	27
II	37	12	19
III	20	43	36
IV	13	19	17
V	—	2	2

Differences in organization are ranked as most important by the English group, and as least significant by the Scottish; differences in way of life, outlook and temperament are so strongly emphasized by the Scottish group as to appear most important when both groups are taken together.

Of those of Bartlett's subjects who predicted breakdown, 33% did so because of the difference in I.Q.; 26% held that breakdown must occur because one group is democratically led and the other not; 20% stressed the conservatism of the Labourers, and 15% predicted a clash of interests owing to the increasing use of machinery. But to the practical group the difference in I.Q. appeared far more important and far more antagonistic to co-operation than to the academic group.

In spite of general agreement as to the probability of breakdown, each group shows a tendency to select some particular set of factors which will ultimately bring it about. This strongly suggests that the selection of specific conditions as favourable or unfavourable to social co-operation is itself specially determined. Further experiment, with different groups, will have to be carried out both with these generalized conditions and with others.

VI. Case III

(i) *Material of Case III: "An anachronism removed"*

The Amateur Rowing Association and the Henley stewards are to be congratulated on their decision to remove the invidious bar which they have hitherto maintained against manual workers. Most oarsmen have long regarded as a blot on the sport the section of the amateur definition which excluded every one "who is or has been, by trade or employment for wages, a mechanic, artisan, or labourer or engaged in any menial duty". So long ago as 1890 a rival body, the National Amateur Rowing Association, was formed on the initiative of such indisputable amateurs as Dr F. J. Furnival, the Reverend Sidney Swann, and Prebendary Propert, with the object of providing opportunities for rowing for amateurs without distinction of occupation. Encouraged by rowing men such as Lord Desborough, Lord Iveagh, and Captain R. C. Bourne, the N.A.R.A., which was reconstituted in 1930, has grown until its affiliated clubs now number about a hundred with some thousands of active oarsmen. None of its crews, however, are eligible to compete at Henley or at any of the A.R.A. regattas, and none of the members of the A.R.A. are allowed by their association to row in regattas organized by the N.A.R.A., although, except for the class bar, the amateur definition of the two associations is in substance the same. This anachronism, reminiscent of obsolete social distinctions, will cease to exist at the end of the present season, when all genuine amateurs, whatever their position in life, will be admitted to compete in all amateur regattas, including Henley, whether held under the auspices of the A.R.A. or of the N.A.R.A.

There can be no doubt that the sport has suffered from the exclusion of a great source of promising material from all possibility of ever taking part in the classic events. The rigid division into sheep and goats who must not be allowed to mix, inevitably interfered with the development of rowing and with the promotion of good fellowship, one of the chief objects of every sport. Cases such as that lamented in a recent letter from the President of the Southampton University Boat Club have not been rare. Rowing men whose memories are long enough will remember the unpleasant effect produced when an N.A.R.A. eight was refused permission to compete for the Cup offered by the King at the regatta held to celebrate the end of the War. In the event this Cup was won by an Australian crew, to whose members the bar against manual workers must have seemed incomprehensible, since no such class distinction would be tolerated in any sport in Australia or any of the other Dominions. One of the worst effects of the rule now rescinded is that it has deterred Dominion crews who have wished to measure themselves against English oarsmen from competing at Henley. When they have been told that their entry could not be accepted because some of them worked for their living with their hands they have been at first incredulous and then contemptuous of what they regarded as another exhibition of English snobbery. In this country, where its origin has been understood, as well as the natural desire to preserve the family party atmosphere characteristic of English regattas, the bar has caused very little ill-feeling. The N.A.R.A. would be the first to acknowledge the sympathy and the help they have received both from the A.R.A. and from the Henley stewards. Now that an outworn distinction has been swept away, there is every prospect of increasing co-operation between the two associations, to the benefit of both and of English rowing as a whole.

This passage appeared as a leading article in *The Times* in June 1937, and is reproduced here by kind permission of the Editor. It was chosen as experimental material primarily in order to check the predictions made on the basis of it against the actual course of events. In general style (being a judicious mixture of historical facts and conventional generalizations) it is eminently suitable for presentation to groups both of experts and of non-experts and for effecting a comparison between them.

(ii) *Method of presentation*

The material was presented to three groups—an academic non-rowing group, a non-academic non-rowing group, and an academic rowing group. The first two groups included both English and Scottish subjects, whose opinions are so similar as not to require separation. The third is a group of Cambridge rowing men. Some variation in the methods of presentation proved necessary. Results were obtained from the first and third groups partly by individual and partly by group presentation, and from the second only by correspondence. To what effect this affected the answers is not known. As before, each subject was allowed to refer to the material as much as he liked in dealing with it. The question asked was:

Do you think that there is every prospect of increasing co-operation between the two associations? State your reasons.

VII. Analysis of results of Case III

(1) *Numerical analysis*

Table XX

Group	Yes	No	Doubtful	Total
I. Non-rowing academic	80	22	18	120
II. Non-rowing non-academic	16	5	9	30
III. Rowing academic	10	2	18	30
Total	106	29	45	180

Table XX gives a simple numerical analysis of the results of Case III, and Table XXI the same results expressed as percentages of the total number of responses.

Table XXI

Group	Yes %	No %	Doubtful %
I	66·7	18·3	15
II	53·3	16·7	30
III	33·3	6·6	60

The non-academic non-rowing group is one of upper working and middle-class persons, and the numerical analysis reveals the variation of opinion between them and the first group. It is, however, rather misleading as regards the third. The first two groups, without technical knowledge, are inclined to accept the phrase 'increasing co-operation' without much analysis, and to give a straightforward affirmative or negative answer. The doubtfuls in each of these groups are subjects who are unwilling to be definite because their knowledge of the facts is extremely limited. In the expert group, however, while the number of direct negatives is remarkably small, the number of doubtfuls is as high as 60%. This is due not to lack of knowledge of relevant data, but to excess of it. The doubting experts give an extremely detailed account of what is likely to happen, and owing to the complexity of the situation as they see it, their answers are strictly neither affirmative nor negative.

(2) *Qualitative analysis*

(i) *Increasing co-operation.* Table XXII shows the causes considered likely to bring about increasing co-operation between the two associations.

Table XXII

Cause	N.R.A. (80)	N.R.N.A. (16)	R.A. (10)	Total (106)
I. Common interest in sport eliminates class-distinctions	39	10	8	57
II. Co-operation to benefit of both associations, of this sport as a whole	19	4	2	25
III. Gradual disappearance of class-distinctions	24	5	7	36
IV. Personal contacts	10	9	2	21
V. Other causes	8	2	6	16
Total	100	30	25	155

Increasing co-operation is predicted in all groups mainly on the basis of the conventional generalization that "sport eliminates class distinctions". The practical (non-rowing) group backs this up closely by expressing confidence in another generalization—that "men individually are not snobs, though in a group they may act as such". What is evidently meant is that when two groups which have previously stood apart from one another are brought together, the personal contacts of individual members of each group will go a long way towards breaking down class distinction, and creating a real understanding.

The disappearance of class barriers throughout this country as a whole is most clearly appreciated as relevant to this case by the two academic groups.

The benefits of co-operation are envisaged in terms of prestige rather than of economic gain, though the latter receives some attention.

(ii) *No real co-operation.* Table XXIII shows the frequency with which each group selects different conditions as being likely to prevent any real co-operation.

Table XXIII

Causes likely to prevent any real co-operation	N.R.A. (22)	N.R.N.A. (5)	R.A. (2)	Total (29)
I. Sheer snobbery	2	—	2	4
II. Incapacity to see each other's point of view	4	—	—	4
III. Rivalry within nominally united body	15	4	—	19
IV. "The peculiar rowing tradition"	2	—	—	2
V. A.R.A. removed bar only in deference to public opinion	2	3	—	5
VI. Nature of the sport of rowing itself (no real analogy with other sports)	—	—	2	2
Total	25	7	4	36

Sheer snobbery on the part of the A.R.A., in the sense that they will refuse to recognize the N.A.R.A. is little anticipated. On the other hand, condescending recognition on the one hand, with self-assertiveness and an inferiority complex on the other, are considered to be the most probable causes of rivalry and jealousy, and of a merely nominal union.

Two rowing men mention that, "although sport is nominally a social link because of the close contact into which sportsmen come, this is not so in rowing. All intercourse is merely social." In other words, the crews train and race as separate, self-contained units, and only meet under the normal conditions of social life. Both the men who hold this opinion also state that "in the elegant atmosphere of Henley the N.A.R.A.'s would be marked men, and their presence would be disagreeable to undergraduates who consider rowing a University monopoly, and try to separate themselves from the working class".

(iii) '*Doubtfuls.*' The rowing group are alone of interest here. A typical answer would run somewhat as follows: "In England there are two distinct standards of A.R.A. rowing—(*a*) Henley and London, (*b*) Northern provincial.

"(*a*) *Henley*—little co-operation likely, either in the sport itself, or socially. Few N.A.R.A. crews (for reasons of time and expense) can hope to compete successfully with University and Tideway (University trained) crews, and so few are likely to row in the regatta. Nor will there be much social co-operation, for even now there is little contact between the various clubs, and one meets only people one knows already, and their friends.

"*Other Thames regattas*—the co-operation depends on the number and rowing strength of the various N.A.R.A. clubs on the Thames.

"(*b*) *Northern regattas*. The A.R.A. and N.A.R.A. clubs should join together very well. All are of the same social class, financial position and ability as oarsmen; and since the same crews compete against each other on several occasions at different regattas, they cannot help getting to know one another, and increasing social intercourse seems highly probable.

"The standard of N.A.R.A. rowing might be improved if A.R.A. men would row for N.A.R.A. clubs, but it is unlikely that they will do so, even though the bar has been lifted."

This eclectic answer is interesting as showing (*a*) that the experts are far less guided by the governing generalizations of the leading article, than are the non-experts, (*b*) that they tend to be much more specific in their predictions, and (*c*) that they introduce a factor which is of primary importance in securing co-operation, and which is omitted from the leading article, and for that reason by all the non-experts, viz. the improvement of the standard of rowing in N.A.R.A. clubs.

VIII. SUMMARY OF RESULTS OF CASES I, II AND III

The analysis of Case I reveals how in dealing with certain group relationships, five selected groups of subjects made either implicit or explicit use of the following main general principles:

(1) (*a*) Groups are seldom, if ever, capable of disinterested conduct, and the co-operation of groups depends on how far the interest of the group is furthered thereby, and the pride of the members in their own group, which flatters their individual self-esteem, satisfied.

(*b*) The causes which facilitate and consolidate the union of religious groups (and probably of most other groups as well) are generally very far removed in tone from so-called Christian motives.

(2) (*a*) A group surcharged with emotion, or one which has just passed through a period of emotional stress, is generally more willing to co-operate than when in its normal state.

(*b*) The co-operation thus secured is apt to break down unless stabilized by other causes.

(3) Most groups display a natural conservative tendency to unite against disruptive factors.

(4) Any two groups which fulfil much the same function in a community and which, while vigorous, resist all attempts at union, will co-operate immediately they are so reduced in numbers or vitality as to be faced with destruction.

(5) The central control which a large organization exercises over its sub-groups facilitates the co-operation of these groups.

(6) Religious differences, being the outcome of inescapable temperamental characteristics, cannot be resolved by any externally imposed union.

(7) (*a*) Country and seafaring folk derive from their way of life a temperament and cast of mind wholly unsuited to consort with an urban population.

(*b*) This difference is more than one of economic organization, but the first step towards securing co-operation between an urban group and a group of country or fisher folk would be a change in economic organization.

(8) *Either* A rebel 'within a tradition' is always supported, but becomes ineffective once he sets himself up as wholly independent.

Or A 'cause' receives support only in so far as, and as long as, it makes an eloquent appeal to a man's natural egotism and flatters his belief in his own importance.

While Case II does not yield any generalizations over and above those selected as leading conditions by the experimenter, it reveals very effectively the importance which is placed upon those generalizations by different social groups. Further experimentation is required to show how the use of those generalizations is itself specially conditioned (p. 334 above). Case II shows also that, while there may be close agreement as to the chances of co-operation or breakdown in any given situation, there may be considerable disagreement as to the precise psychological conditions which bring about co-operation or breakdown.

Case III is interesting chiefly as illustrating how 'psychological remoteness' determines the manner in which subjects deal with a given situation (see below, p. 343).

The accuracy of the predictions made by the two groups of non-expert subjects, and by the group of experts, will in due course be checked against the actual course of events.

IX. Discussion

This paper has been written primarily to illustrate a method. I shall therefore conclude with some general remarks on the method, with special reference to our wider scheme of research, and with a view to meeting possible criticisms.

(1) *Material*

(i) *The questionnaire.* The method has, for our purpose at least, two advantages over most questionnaires: (1) By referring to a particular concrete case, the experimenter can be reasonably sure that his questions have the same essential meaning for his subjects as they have for himself. (2) If we ask an individual a set of isolated questions concerning the nature and relative importance of the factors governing group relationships, then we can scarcely hope to know what aspects of his experience

he has recourse to in framing his answers. If, however, we describe a situation in which certain postulated factors are *possibly* relevant, then we can observe what selection and weighting of these factors occurs, and what further factors, if any, are imported. By thus controlling the types of situation used as frames of reference, and comparing the response of different social groups to different situations, we may perhaps discover how this all-important selection and weighting of governing conditions is itself specially determined (cf. p. 334 above).

(ii) *Detailed and generalized descriptions.* The experiments just described, however, only justify us in asking what important differences are found between those cases in which the conditions of possible co-operation have to be inferred from a considerable amount of particular detail (Case I, p. 210) and those in which they are presented in a generalized form (Case II, p. 330 and Case III, p. 335). The differences are as follows:

(1) A quick return can best be obtained by the use of examples similar to Case II. The issue being stated in terms of a few generalizations, most subjects solve it simply by utilizing these, and their opinions can be quickly classified. By the further use of similar material it should be possible speedily to determine the frequency and distribution of *any* conventionalized generalization throughout a given social group.

In so far as widespread opinion (either *true* or *false*) concerning the factors governing social co-operation affects the chances of co-operation, an enquiry into the frequency and distribution of such opinion will have practical value.

(2) Very detailed descriptions such as those of Case I, on the other hand, yield results which require more careful sifting. This is chiefly because, as noted above (p. 220), the psychological principles on which many solutions are based are frequently not explicitly stated, but have to be inferred and formulated by the experimenter.

Case I (unlike Cases II and III) yielded in addition to these unformulated principles a number of unsolicited generalizations.

Very detailed descriptions thus seem to have a double value. Firstly, they reveal the presence of conventionalized generalizations the frequency and distribution of which can be more fully determined in the manner described above—i.e. by embodying them in less detailed descriptions, and presenting them to selected groups of subjects. Secondly, when subjects are asked to state the probable upshot of a situation described in great detail, some of them, avoiding easy generalizations, base their opinion largely on specific conditions selected from the given context.

Underlying the specific conditions so selected it may sometimes (as in Case I above, pp. 220, 221 and 222f.) be possible to discover a principle of fundamental importance. This principle may never have been consciously formulated as a principle by any of the subjects in question, yet in so far as they make implicit or unconscious use of it, we can infer that they have had, in real life, empirical evidence of its working and are able to recognize factors relevant to it. A more extended use of material similar to that of Case I should therefore enable us profitably to tap the actual working experience of individuals within selected social groups and so to gain some indication of the *real* factors which govern social harmony and discord.

To sum up: (1) generalized descriptions are most suitable for showing how people think about social questions (p. 340); (2) very detailed descriptions are most likely to provide us with the guiding principles essential for profitable observations and experimentation in the field (p. 208, ii).

(2) *Some possible criticisms*

(i) *Concreteness.* It will probably be urged that in spite of its claim to concreteness our material must, by its very nature, remain viciously abstract. All descriptions being essentially the result of selection and interpretation, no description can ever represent an actual situation in its full complexity. In asking our subjects to deal with described situations, we are therefore asking them to deal with biased and incomplete pictures of the actual.

The reply to this is twofold: (1) In dealing with an actual situation in which a practical decision has to be made, every person has to exercise some kind of selection. He cannot hope to know *every* aspect of the situation, and from those of which he is aware, he has to pick out the relevant from the irrelevant. In other words, the issue is decided, not in terms of the whole situation but in terms of a few selected factors. In the same way, when we are contemplating others acting in a real situation, we must, if we wish to predict the probable course of events, select those factors which we consider will govern the agents' actions. (2) In confronting our subjects with a described situation, and asking them to predict how they themselves or others would act under those conditions, we are merely presenting them with a case in which some of this essential work of selection has already been carried out. As we have seen (p. 340), it is, for the purposes of our enquiry, a definite advantage to be able thus to keep a check on our subjects' summing-up of the situation, and to observe any further selection or importation of data which may deter-

mine their predictions. The risk of suggesting the probable issue has simply to be taken.[1]

(ii) *Psychological remoteness.* It may be objected, again, that there is little profit in collecting opinions from subjects concerning situations from which they are 'psychologically remote'. In this context, a subject would be psychologically remote from any situation which describes group-relationships of which he has had little or no experience either as an observer or as a participant.

The experiments already carried out (most notably Case III, above) indicate that most individuals show remarkable confidence in dealing with situations which are psychologically remote in this sense. As we saw, in predicting the probable course of events, they readily fall back on conventionalized generalizations concerning group relationships. The practical value of determining the nature, frequency and distribution of these generalizations may be disputed, but an enquiry into the extent to which such generalizations govern the way in which people think about social questions would be of undoubted theoretical interest. A beginning at one end of the scale has already been made by collecting the opinion of a number of subjects concerning a group relationship in which they are actual participants. The investigation will be continued with material of ever-increasing remoteness, and a report will be published in due course.

(iii) *Real and described situations.* Finally, it may be maintained that *every* described situation, just because it is described, must remain psychologically remote. No matter how closely a described situation may concern us, and no matter how vivid our description, the description can never possess the force and urgency of the actual; and therefore our reaction to the description can give no sure indication of our reaction to the real situation. Implying as it does, that we have neither knowledge nor control of the key factors which govern our behaviour as members of a group, this criticism raises issues which it is impossible to discuss here.

[1] Cf. H. Cantril, "The prediction of social events", *J. Abn. Soc. Psychol.* (1938), XXXVI: "The uniformity of predictive judgements regarding a social event is in part, but not wholly, proportional to the degree to which the specific factors leading up to an event can be resolved into a consistent and meaningful pattern." Cantril reaches this conclusion after using the *questionnaire* method to obtain, from selected individuals belonging to eleven different social groups, their opinion as to the probable upshot of a variety of contemporary situations familiar to an informed American. Just how far a deliberate resolving of these situations by the experimenter into "consistent and meaningful patterns" biases subjects' predictions could be at least partly determined by embodying Cantril's (or other) questions in descriptions such as Case III above, and using them with similar groups of subjects.

For the present we can only point out that the criticism, if pushed to extremes, would apply to *all* investigation in psychology which uses situations built up under experimental conditions; and, once again, that this particular investigation is a preliminary one, expressly devised to give direction to, and be supplemented by, other methods.

X. Summary

I. Three interdependent lines of research are proposed:

(*a*) Direct observation of limited communities.

(*b*) Controlled experiments in the field.

(*c*) Experiments in constructive social thinking.

The conclusions which Bartlett reached after a series of experiments of the last type are summarized.

II. Three further experiments in constructive social thinking are described, and their results analysed.

III. An attempt is made to assess the significance of these results both in themselves and in relation to the wider scheme of research. The discussion therefore is concerned mainly with the theoretical validity and the practical value of the method of investigation employed.

I have to thank Prof. F. C. Bartlett for his inspiration and help; and Prof. James Drever for his kindness in allowing me to conduct group experiments in his laboratory in Edinburgh.

(*Manuscript received* 15 *April* 1938)

VARIABILITY OF JUDGEMENT AND 'STEADINESS' OF CHARACTER

By JEAN D. CUMMINGS

I. *Introduction* (pp. 346–347).

II. *The experiment* (pp. 347–350).
- (1) *The block-in scale* (A) (p. 347).
- (2) *The hypothetical situation scale* (B) (pp. 347–348).
- (3) *The Marston scale of introversion-extraversion* (p. 348).
- (4) *The tests* (pp. 348–350).
- (5) *Subjects' introspections* (p. 350).

III. *The test results* (pp. 350–359).
- (1) *Reliability* (pp. 350–352).
- (2) *Comparison of continuous and day-to-day work* (pp. 352–353).
- (3) *Intercorrelations of error, speed and variability* (pp. 353–355).
- (4) *Intercorrelation of the tests* (pp. 355–359).

IV. *Rating scale results* (pp. 359–363).
- (1) *The block-in scale* (A) (pp. 359–361).
- (2) *The hypothetical situation scale* (B) (pp. 361–362).
- (3) *The Marston introversion-extraversion scale* (p. 362).
- (4) *Reliability of the scales* (p. 362).
- (5) *Intercorrelation of rating scale results* (pp. 362–363).

V. *Correlation of rating scales and test results* (pp. 363–367).
- (1) *Ratings and variability scores* (pp. 363–366).
- (2) *Ratings and error scores* (p. 366).
- (3) *Ratings and speed* (p. 366).
- (4) *Correlation of variability with the whole block-in scale* (pp. 366–367).

VI. *Summary of main results* (pp. 367–368).

References (pp. 368–369).

I. Introduction

Of the problems of vocational selection, one of the most difficult seems to be that of forecasting the 'reliability' or 'conscientiousness' of an individual and his habits of work. The measurement of his intelligence and special abilities gives no measure of his tendency to work conscientiously and thoroughly at a monotonous task, or at a task calling for concentrated attention to detail.

In the present experiment it was thought that the '*w*' factor of Webb's(1) research, which he described as being in close relation to 'persistence of motives', might be revealed in the various features of the work curve for a group of subjects. In particular, it was thought that its relation to the following measures might be of interest:

.(1) The average error of judgement in each 3 min. period for 20 periods.

(2) The average variation of judgement in each 3 min. period for 20 periods.

(3) The average speed of judgement in each 3 min. period for 20 periods.

With regard to (2) above, Cox(2) has shown that in tests of manual skill, the daily variability of his more able subjects was less than that of his less able subjects. Walton(3), in short-period work with children, found that small amplitude of oscillation in two tests of motor abilities and two of cognitive processes correlated with estimates of 'steadiness' of character. There is therefore ground for enquiring whether variability (in this case variability of judgement) may not be correlated with low 'steadiness', i.e. correlated negatively with '*w*'.

The analysis of oscillation has shown '*o*' to be a general and not a specific factor, but it cannot be assumed that variability in more complex tests is anything more than a characteristic specific to each test, caused by skill, interest and other factors. Hollingworth(4) could produce no conclusive evidence for a variability factor in such tasks as tapping, colour naming and in the variations of the pulse rate, although Asch(5), repeating the experiment, but eliminating the pulse-rate scores, found a systematic though slight factor common to the tests (average r was 0·1972).

With regard to measures of 'persistence', tests so far devised seem to have shown that an individual's persistence score is specific to the task he does. May & Hartshorne(6) have shown this conclusively in short-period tests. But there is a great deal of difference in the usage of the

term 'persistence'. May and Hartshorne use it to mean "the ability to resist the tendency to go more and more slowly". Crutcher(7) uses it to mean the tendency to continue definite activities in the face of difficulties. Howells(8) measures it by the amount of endurance with which a person can bear increasing discomfort and fatigue. One of the chief difficulties seems to be that in miniature situations the situation is not alike—does not mean the same—for every subject.

One aim of this experiment was to see, by independent estimates, whether several uses of the word 'persistence' do in fact denote a group factor—whether for instance 'persistence in routine', 'persistence of interests', 'persistence of principles of conduct', etc., are highly intercorrelated: finally, given any reliable estimate of persistence, to see whether this was revealed in any feature of the work curve.

II. THE EXPERIMENT

The subjects were eighteen women undergraduates in their second year at college. Each subject was asked to obtain three judges from among intimate friends, each of whom would assess her in the rating scales; and to assess herself on the same scales independently.

The procedure of testing has been previously described in more detail(9), but was, briefly, as follows:

(1) *The block-in scale* (A)

In this scale the rater was asked to assess the subject in thirty traits, selected as '*w*' traits, plus additional traits relating more closely to work qualities and persistence in general. The traits were listed on the left-hand side of the page and on the right was a series of blanks, which the rater could block in to any degree she thought correct. Thus:

	Very low	Very high
Permanence of interests	////////////	

No attempt was made to control the distribution of the values because only a rank order was desired.

(2) *The hypothetical situation scale* (B)

In this scale the rater was given the following instructions:

"Imagine the subject is placed in the following situations and put a tick opposite the action or actions you think she would be most likely to take. If you are unsure, or if you have anything to add, put it in the space left for REMARKS."

The five questions were designed to measure by particular examples:

Persistence in routine work (checking figures).
Persistence of principles (steadfast support of a society).
Persistence of interests (length of time hobbies last).
Persistence of intentions (changing holiday plans).
Persistence of moods (persistently cheerful or despondent).

The questions correspond to the five persistence traits in scale A, but in this scale B it was thought desirable to describe the traits in specific situations with which anyone might be faced.

(3) *The Marston scale of introversion-extraversion*

A measure of introversion was obtained by using the Marston scale, which although criticized by Guilford & Hunt(10) as being based on the emotional aspect of the introversion-extraversion distinction, was thought to be sufficiently satisfactory to obtain a rank order in a research upon only a small number of subjects.

(4) *The tests*

Testing was individual and was carried on for 5 days a week for 3 weeks, Saturdays and Sundays being omitted. On the first day there was a 3 min. practice period in each of the four tests. On each of the other 4 days of the first week the subjects did continuous work for half an hour, doing one test on each day. In the remaining 2 weeks they did 3 min. on each test each day, making a 12 min. testing time each day. The total testing time for each subject was 4 hr., neglecting of course the time taken between the tests. The tests were:

The arm-movement test (I). The material for this test consisted of two boards, each with two parallel grooves along which small square blocks could be moved and the distance registered in millimetres. Clamps could be inserted at any point in the scale to limit the distance the blocks could be moved. Two such boards were placed before the subject, so that there were four parallel grooves. The subject, starting with the topmost groove, held two small blocks on either side of a central clamp, and with her left hand ran a block as far as it would go away from the clamp. She then ran the right-hand block in the opposite direction to what she considered was an equal distance. She made the judgement with closed eyes, opening them between judgements to see the amount of error she had made. She ran the blocks along the lower grooves in the same way, and then immediately returned to the topmost, and so on without interruption. She

was instructed to choose her own speed of work, which would not influence her score in any way. The speed of each judgement was, however, recorded. The length of the standard distance, i.e. the left-hand run, was varied slightly in each case, so that a separate judgement had to be made for each groove.

The ball-and-slot test (II). The material in this case consisted of a series of boxes; the subject placed a ball in the top of the box, allowed it to run along the top of a metal ledge and from there drop into one of twelve slots, marked from 1 to 12. The subject tried in every case to drop the ball into number 6 slot, although the preliminary run along the metal ledge made it more difficult to do this than to drop the ball into numbers 1 to 5. The subject worked, as in test I, without interruption and at what speed she chose. The only restriction was that the box must be held at an angle of 45° or more with the horizontal.

The character judgement test (III). The material used in this test was the same as for the block-in rating scale described above. In the practice given before the test, the subject assessed herself in all the traits and blocked in one complete sheet, so that in the test proper there was no great difference in speed between the first set of judgements and subsequent ones, but only a gradual quickening throughout the test. The subject worked continuously through the sheets, each of which contained the same thirty character traits in which self-assessment was to be made. The traits were, however, arranged in four different orders to prevent the subject remembering the general look of a completed sheet, and in order that judgement had to be made on each trait separately.

Matching greys test (IV). In this test the subject was given a board in which two small squares were cut out. Through the left of these a standard grey was shown, while on the right a revolving disk at the back of the board displayed a series of greys ranging from a grey easily discernible as lighter than the standard to one very much darker. The subject had to judge which one of this series most nearly resembled the standard. Four standards were used in turn, varying slightly in brightness. The subject started from the light and dark ends of the scale alternately, but could reverse the disk as often as she pleased.

From these tests measures were obtained of:

(*a*) The amount of error of each judgement.

(*b*) The time of each judgement.

(*c*) The variability of each 3 min. period. This was measured in tests I and II by the average deviation from the mean of the judgements (neglecting the error). In tests III and IV the variability was obtained by

subtracting each judgement from the previous corresponding one and averaging these subtractions regardless of sign for each 3 min. period.

(*d*) The total variability of each subject was obtained for the continuous and subsequent day-to-day work by averaging the variabilities of the 3 min. periods (ten for each type of work).

(5) *Subjects' introspections*

Subjects were asked for introspections in general, and they were also asked:

Which test was the most interesting?
Which test was the most uninteresting?
Which test needed most concentration?
Which test was most physically fatiguing?
In which test was the subject most successful?

III. The test results

(1) *Reliability*

The reliability of the variability scores was calculated in two ways—by correlating the variability scores on odd periods with those obtained on even periods, and by correlating scores on the first five periods with those on the last five periods in both types of work.

The reliability coefficients are shown in Table I, the figures in brackets being the coefficients corrected by the Spearman-Brown prophecy formula to give the reliability of the whole test.

Table I. *Variability*

	Test I		Test II		Test III		Test IV	
	A	B	A	B	A	B	A	B
Odd and even judgements	0·71 (0·83)	0·77 (0·87)	0·76 (0·86)	0·86 (0·93)	0·88 (0·94)	0·81 (0·89)	0·03 (0·06)	0·79 (0·88)
First 5 and last 5 periods	0·78 (0·88)	0·69 (0·82)	0·74 (0·85)	0·83 (0·91)	0·70 (0·82)	0·70 (0·82)	0·14 (0·25)	0·38 (0·55)

It will be seen from the table that while the reliability of tests I, II and III is satisfactory, that of test IV is markedly not so, the reliability of odd and even judgements in the continuous work (A) being as low as 0·06. There are several possible reasons for this low reliability. The subjects from their introspections evidently detected qualitative differences in the grey series, which led them to judge as equal greys which were several shades apart in brightness, but which were nevertheless more 'blue' than the others, and so for this discrepant reason similar. The sub-

ject sat always in the same position in the room and in artificial light, but other conditions such as eye fatigue from previous work, or the brightness of light outside the test room, may have altered the day-to-day results. Whatever the reason, it seemed obvious that the test was not sufficiently reliable to justify its use, and it was accordingly discarded.

The other three tests are all of good reliability. All the corrected coefficients are above 0·80, and it seems therefore that both the continuous and day-to-day results are reliable and almost equally so. This finding is a somewhat surprising one, especially in the block-in scale of character judgements, for it is difficult to see why a subject should go through twenty 3 min. periods altering her character judgements and speeding up the rate of making them but in variability should keep so constant a position relative to other subjects that the rank orders at the middle and end of testing should correspond so highly. It would perhaps be easier to understand if there had been a wide range of scores, but the whole length of the block-in scale was only 5 cm., and the range of average scores only 2·57 mm. Why, for instance, should one subject have 3 min. variability scores of 0·185, 0·145, 0·112, 0·172, etc., while another subject scores 0·240, 0·220, 0·216, 0·220, etc.? How can this difference, which is a difference between two series of individual scores, be accounted for?

The reliability of the error scores (there was of course no error scoring in the character judgement test) is shown in Table II. The reliability of

Table II. *Error scores*

	Test I		Test II		Test IV	
	A	B	A	B	A	B
Odd and even judgements	0·89 (0·94)	0·94 (0·97)	0·96 (0·98)	0·95 (0·97)	0·60 (0·75)	0·24 (0·39)
First 5 and last 5 periods	0·63 (0·77)	0·89 (0·94)	0·87 (0·93)	0·89 (0·94)	0·21 (0·35)	0·58 (0·73)

test IV (grey matching) is again unsatisfactory, and this test was finally discarded in working out the results. The reliability of the other two tests is satisfactory.

The reliability of the speed scores is shown in Table III. With one exception the coefficients suggest an exceedingly consistent rate of work, consistent both from period to period and between the beginning and end of work. The exception is the continuous work periods of the character judgement test (III). Here the coefficient of 0·68 indicates that although the times of judgement are highly consistent from period to period, there

Table III. *Speed scores*

	Test I		Test II		Test III	
	A	B	A	B	A	B
Odd and even judgements	0·99 (0·99)	0·97 (0·98)	0·94 (0·97)	0·98 (0·99)	0·91 (0·95)	0·97 (0·98)
First 5 and last 5 periods	0·89 (0·94)	0·75 (0·86)	0·83 (0·91)	0·81 (0·89)	0·51 (0·68)	0·79 (0·88)

are individual differences in speeding up or slowing down which affect the rank order when the first and last halves of the test are compared; it is of course not surprising that individuals differed in the way they increased their speed of self-judgement. By the time the day-to-day work was reached individual rhythm of work was evident and remained unaltered except for a slight increase of speed common to all.

It thus appears that all three measures—speed, error and variability —are highly consistent within each of the three activities of the tests. That is to say, a subject tends, relative to other subjects, to keep speed, error and variability consistent throughout one half-hour of continuous work and throughout the following 10 days of daily testing. Differences in improvability and speeding up may alter this relationship somewhat, but in these particular tests their effect is not marked.

(2) *Comparison of continuous and day-to-day work*

Table IV gives the correlations obtained between the continuous half-hour work (A) and the daily 3 min. period work (B), the variability scores

Table IV. *Variability scores*

		Continuous work (A)		
	Test	I	II	III
Day-to-day work (B)	I	0·36		
	II		0·90	
	III			0·80

being considered first. Two of these correlations are significant, showing that in the ball-and-slot and character judgement tests a subject's tendency to vary is very much the same in uninterrupted work and in work spread over 2 weeks. In the arm-movement test (I) there is no such close relationship, and although there are several possible reasons for this they are merely speculative, and the cause cannot be discovered without much more information upon the rate of learning in this test, conditions of boredom, fatigue, etc.

This correlation of variability scores in continuous and in day-to-day work is important chiefly from the point of view of the framing of a diagnostic 'variability' test. If test material can be obtained in which a subject's variability can be assessed as well in an initial long period as in periods spread over a length of time, such material is obviously of more practical use for testing purposes than material which reveals a subject's tendency to variation only after a length of time. From this point of view, the ball-and-slot and character judgement tests are better than the arm-movement test, in which it is quite possible the A type of work is merely an 'absorption' period, i.e. a period into which many extraneous factors enter, to be gradually levelled out and their attenuating influence decreased. The ball-and-slot and character judgement tests are reliable both in long and short period work; a subject's scores on both types of work are highly consistent, and in fact his 'variability' could be diagnosed in one half-hour test with no need for a longer period.

Table V. *Error scores*

		Continuous work (A)	
	Test	I	II
Day-to-day work (B)	I	0·71	
	II		0·93

Table VI. *Speed scores*

		Continuous work (A)		
	Test	I	II	III
Day-to-day work (B)	I	0·63		
	II		0·77	
	III			0·62

The correlations for the error and speed scores are given in Tables V and VI. These correlations are lower, but significant, showing that there is a tendency for an individual's rhythm of work on a test to be consistent whether that work is over a long or short period.

(3) *Intercorrelations of error, speed and variability*

The intercorrelations of error and variability are given in Table VII. It seems evident that the correlation of error scores with variability depends on the nature of the test. In the continuous work of test I highly variable subjects were not necessarily the most accurate. In test II on the other hand the correspondence was from the beginning very high.

Table VII

			Variability I		Variability II	
			A	B	A	B
Error	Test I	A	0·27			
		B		0·73		
	Test II	A			0·91	
		B				0·94

It is easily understood from the nature of the ball-and-slot test that a subject who was skilful in getting balls into slot 6 would vary less than one who was unsuccessful. A close correspondence of error and variability probably depends therefore upon the nature of the test material, in which success may imply lack of variation (as in the ball-and-slot) or be comparatively independent of it. It must also, evidently, depend upon the degree of knowledge of success given to the subject and the incentive to vary upon knowledge of failure.

The variability and speed correlations were taken between low variability and low speed and were obtained for the first five A periods (A_1), the last five A periods (A_2), the first five B periods (B_1) and the last five B periods (B_2). The results, shown in Table VIII, should therefore show

Table VIII

Variability

Test I (arm-movement)

Speed	A_1	A_2	A	B_1	B_2	B
A_1	−0·15					
A_2		−0·07				
A			*−0·08*			
B_1				−0·41		
B_2					−0·39	
B						*−0·45*

Test II (ball-and-slot)

Speed	A_1	A_2	A	B_1	B_2	B
A_1	0·53					
A_2		0·42				
A			*0·51*			
B_1				−0·10		
B_2					0·22	
B						*0·06*

Test III (character judgement)

Speed	A_1	A_2	A	B_1	B_2	B
A_1	0·05					
A_2		0·03				
A			*−0·14*			
B_1				−0·14		
B_2					0·48	
B						*0·06*

how the relationship of variability and speed altered over the 3 weeks' testing time. The italicized figures give the correlations for the whole A and B periods. It seems that the relationship of variability to speed depends upon the nature of the test. In test I there seems to be a slight tendency, more marked in the later testing, for speedy subjects to be the least variable, but this relationship is inverted in the case of the ball-and-slot test; whereas in test III there seems to be no significant relationship.

The error and speed correlations are given in Table IX. Here again, no decisive trend is to be found. Lack of error correlates negatively with slowness in test I, indicating that a fairly high speed is conducive both to lack of error and variability, but this is not the case in test II.

Table IX

Error

Speed	Test I						Test II					
	A_1	A_2	A	B_1	B_2	B	A_1	A_2	A	B_1	B_2	B
A_1	−0·05						0·06					
A_2		−0·01						0·35				
A			*−0·24*						*0·25*			
B_1				−0·44						−0·30		
B_2					−0·26						0·10	
B						*−0·37*						*−0·04*

The relations of speed, variability and error seem, therefore, to be dependent on the nature of the test and the progress of it; there is no evidence that any relationship holds between them that is not specific to each test.

(4) *Intercorrelation of the tests*

The question now arises as to whether these measures of variability, error and speed are specific to each test, or are wider factors common to all three tests. The variability results have been correlated at the end of each five periods as before, i.e. at A_1, A_2, B_1 and B_2 as shown in Table X. The coefficients for the whole A and B periods are italicized.

Table X. *Variability scores*

		Test I (arm-movement)						Test II (ball-and-slot)					
		A_1	A_2	A	B_1	B_2	B	A_1	A_2	A	B_1	B_2	B
Test II (ball-and-slot)	A_1	0·32											
	A_2		0·20										
	A			*0·19*									
	B_1				0·60								
	B_2					0·61							
	B						*0·81*						
Test III (character judgement)	A_1	−0·05						0·21					
	A_2		0·12						0·39				
	A			*0·04*						*0·34*			
	B_1				0·20						0·42		
	B_2					0·27						0·25	
	B						*0·37*						*0·47*

The coefficients in this table, although not all significant, are, with one exception, all positive, i.e. correlations taken at four different points in the progress of the test yield positive coefficients throughout. This means that there is a tendency for a subject who is variable at the beginning of test I to be variable also at the beginning of test II and (a little later) in test III, and that throughout the three tests this tendency persists. These correlations, and the fact that on the rankings for the whole B period the average coefficient is 0·55, indicate strongly that in the present sample at least there is either a common factor of variability present; or other extraneous factors, such as overlapping material, are present to account for these coefficients.

The high correlation of 0·81 between tests I and II suggests that possibly the tests are so similar that they are in fact demanding much the same performance, or are at any rate overlapping to a large extent. On the surface there seems to be little in common between the two tests except that they both require good muscular co-ordination and control of movement. A possible group factor is that of 'manual skill', which may underlie both activities. Then a subject who is skilful in one activity will tend to be so in the other, and since variability is highly correlated with error, the chief reason for the high correlation of variability scores will lie in the manual factor. On the other hand, this will hardly explain the correlation with the character judgement test, in which a subject has to judge the extent to which she possesses a quality and block in a scale to that degree.

To settle the question of overlapping material, the correlation of variability scores of the arm-movement and ball-and-slot tests has been

worked out with the error scores partialled out. To do this, the error results of the two tests were pooled (which the degree of correlation justified) and this pooled ranking was partialled out, giving partial coefficients of variability on the two first tests as shown in Table XI.

Table XI. *Variability—error partialled out*

		Test I (arm-movement)					
		A_1	A_2	A	B_1	B_2	B
Test II (ball-and-slot)	A_1	0·45					
	A_2		−0·22				
	A			*−0·27*			
	B_1				0·24		
	B_2					0·25	
	B						*0·37*

In this table, the correlations of the A period (0·45 and −0·22) are accounted for by the fact that in the first half the correlation of error and variability is low while in the second half it is high. Thus the variability correlations of 0·32 and 0·20, when error is partialled out, take on very divergent values because of the difference in the degree of association with the error scores.

The striking feature of this table is that, when error is partialled out between tests I and II, the coefficients come to resemble in magnitude the intercorrelations with test III in which there can be no such influence of error and skill. The results of the A periods become similarly of doubtful significance; the B results show the same low but consistent coefficients.

The conclusion to be drawn from this seems to be that some factor such as manual skill affects both the arm-movement and ball-and-slot tests and their variability scores, and that this makes these variability correlations spuriously high; that when its influence is removed, the intercorrelations of all three tests do indicate the presence of a group factor of variability manifested in the last part of the test—namely, after one half-hour work period has elapsed and the subjects are working on each test for 3 min. a day; that, however, such evidence is only tentative in view of the small number of subjects and in the absence of factor analysis.

The error intercorrelations are given in Table XII. This considerable degree of correlation suggests that throughout these two tests there is some common factor that induces success or failure in both. It is not an extremely strong tendency but it undeniably exists; and it is this that has made the variability intercorrelations artificially high.

Table XII. *Error scores*

Test II (ball-and-slot) \ Test I (arm-movement)	A_1	A_2	A	B_1	B_2	B
A_1	0·53					
A_2		0·50				
A			*0·65*			
B_1				0·59		
B_2					0·57	
B						*0·60*

The intercorrelations of the speed scores are given in Table XIII. These results show that there is no general speed factor operative in these tests. This disagrees with casual observation made of subjects during the tests that some tended to work on all tests more quickly than others. It is in agreement with Vernon's(11) finding in that although there is no general speed factor common to all the tests, there is a high degree of constancy of specific speeds; i.e. subjects keep a very consistent speed in each test, as is shown by the reliability coefficients, although there is no carry-over to other tests.

Table XIII. *Speed scores*

		Test II (ball-and-slot)						Test III (character judgement)					
		A_1	A_2	A	B_1	B_2	B	A_1	A_2	A	B_1	B_2	B
Test I (arm-movement)	A_1	0·01						0·35					
	A_2		0·05						0·07				
	A			*0·00*						*0·32*			
	B_1				0·46						−0·30		
	B_2					0·65						0·16	
	B						*0·66*						*0·01*
Test II (ball-and-slot)	A_1							−0·09					
	A_2								−0·08				
	A									*−0·09*			
	B_1										0·00		
	B_2											0·16	
	B												*0·02*

Variability has here been found to be a much more valuable source of enquiry than the other two measures taken from the work results. It has been found that subjects who are variable in one test tend to be variable in two very different tests, but only after a testing time of half an hour has elapsed. The question immediately arises, how wide is this factor? Does it extend from purely sensory judgements to judgements involving intense concentration or co-ordination of action, or are there different levels or 'fields' of variability? It would be necessary to have a larger

number of subjects to answer these questions, so that each coefficient would be significant in itself, and the correspondence of variability scores in a battery of tests traced more accurately throughout the testing time.

IV. RATING SCALE RESULTS

(1) *The block-in scale* (A)

Traits in this scale were, for correlation purposes, divided into two groups:

'*Persistence' traits:* those traits giving the theoretical aspect of the specific situations of the hypothetical situation scale B;

'*w' traits:* traits of a wider classification called, provisionally, '*w*'. Two of these traits were included in the 'persistence' list and their overlapping has of course to be accounted for.

The intercorrelations of the 'persistence' traits are shown in Table XIV. In this table the correlations are nearly all positive, but 'tendency to planned action' has only a negligible correlation with other traits judged by its average correlation, and in no case does it correlate more highly than 0·31.

Table XIV

Trait		17	24	14	11	Average
Tendency to planned action	(9)	0·27	0·31	−0·11	0·29	0·19
Permanence of interests	(17)		0·22	0·48	0·36	0·33
Thoroughness in routine work	(24)			0·36	0·50	0·35
Strict adhesion to principles	(14)				0·38	0·28
General stability of moods	(11)					0·38

The positive trend of correlation might be accounted for by the existence of a group factor of persistence or of a wider character factor such as '*w*', which would include the persistence qualities. Such factor analysis has not been attempted here, the correlations being used simply as a justification for pooling the scores on those five persistence traits and obtaining a single 'persistence' ranking. This can then be compared and correlated with the ranking obtained on the other measure of persistence, the hypothetical situation scale.

For '*w*' traits, the coefficients given in Table XV were found. It is seen immediately that trait number 27 (depth of thought) correlates negligibly with the other traits; it is hard to say why this should be, except that perhaps 'depth of thought' conveys a different meaning to the reader from 'profoundness of thought' used by Webb. In any case the coefficients are so low that this trait has been excluded, and Table XVI

Table XV

		'*w*' traits							
	Trait	12	13	14	24	27	28	29	Av.
Strength of will	(12)		0·14	0·33	0·15	0·22	0·27	0·35	0·24
Conventionality	(13)			0·31	0·45	−0·10	−0·07	0·34	0·18
Strict adhesion to principles	(14)				0·36	0·02	0·36	0·49	0·31
Thoroughness in routine work	(24)					−0·19	0·69	0·52	0·33
Depth of thought	(27)						−0·01	−0·08	−0·02
Conscientiousness	(28)							0·70	0·32
Trustworthiness	(29)								0·39

Table XVI

Trait	12	13	14	24	28	29	Average
12		0·14	0·33	0·15	0·27	0·35	0·25
13			0·31	0·45	−0·07	0·34	0·23
14				0·36	0·36	0·49	0·37
24					0·69	0·52	0·43
28						0·70	0·39
29							0·48

Average $r = 0{\cdot}36$.

obtained. Here with one exception the coefficients are all positive, the average correlation of 0·36 being slightly higher than the average correlation for 'persistence'.

The uniformity of these coefficients allows a single '*w*' score to be made by pooling the six traits; the term '*w*' is used tentatively, because no attempt at the analysis of '*w*' has been made, the evidence from previous work on the factor being taken for granted. The intercorrelating traits bear sufficient resemblance, it is thought, to the traits described by Webb(1), Garnett(12) and others to justify the final score obtained from them being called a '*w*' score.

The correlation of 'persistence' or '*P*' (this symbol is not used to denote a general or group factor but is merely an abbreviation) with '*w*' score = 0·77. This, however, contains a large element of spurious correlation, since two of the traits were included in both the '*P*' and '*w*' scores. If this spurious correlation is taken as $= 2/\sqrt{30}$ (see Brown & Thomson(13)), there remains a correlation of 0·41 over and above that to be expected from the presence of the common elements. It is therefore possible that since the '*P*' traits correlate rather less among themselves than with the '*w*' scale, their intercorrelation may be due simply to the presence in all of them of some degree of '*w*' factor. This can be neither proved nor disproved in this experiment, but is merely suggested. There is certainly no direct

evidence that these traits of persistence are anything more than '*w*' traits; that, in fact, the particular aspects of persistent behaviour taken here may be manifestations of 'perseverance' or 'will' or whatever else this general factor may be called.

The two scores, '*P*' and '*w*', have, however, been kept distinct for correlation with the other rating scales and finally with the test results.

(2) *The hypothetical situation scale* (B)

Contingency coefficients were obtained between three of the five questions of the scale from a 5×5 table, giving Table XVII. There is thus

Table XVII

Question		1	2	4
Persistence in routine (checking additions)	(1)		0·675	0·636
Permanence of interests (hobbies)	(2)			0·710
Persistence of intentions (keeping to holiday plans)	(4)			—

Average $C = 0{\cdot}674$ (approx. 4 times the P.E.).

fair agreement between situations designed as example situations of 'persistence in routine', 'permanence of interests' and 'persistence of intentions'. At least we are justified in saying that in these eighteen cases there is a decided tendency to judge the person who would be careful at checking additions to be also persistently interested in any hobby she might take up, and again persistent in carrying through her holiday plans: and, on the other hand, the person who would neglect checking precautions would also be the person to lose enthusiasm quickly in a hobby, and to be changeable and impulsive over holiday arrangements.

Between the other two questions, in which the grading was coarser, contingency coefficients on a 2×2 table were as in Table XVIII. There

Table XVIII

Question		3	5	1, 2, 4 pooled
Persistence of principles	(3)		0·330	0·219
Persistence of moods	(5)			0·00

seems to be little if any relation between persistence of principles and persistence of moods, and between these and the other pooled questions of the scale; at least, there is little relation between the tendency to support

societies with whose aims one is in sympathy and persistence of cheerful and despondent moods. On the other hand, lack of correlation may be due to the coarseness of the grading on these two questions.

There is thus some evidence—not conclusive—that persistence is not entirely specific to every activity. At least, to the minds of the judges there is a correspondence between persistence in routine work, in hobbies and in making plans. Webb's term 'perseverance in the face of obstacles' is perhaps another way of expressing this characteristic, since the example situations imply at least some resistance to impulse if not to external obstacles.

(3) *The Marston introversion-extraversion scale*

The subjects were marked on a five-point scale for each question, '1' denoting the extreme introvert score, and '5' the extreme extravert.

The total score for each rater was the sum of the marks of the twenty questions, and the final introversion score was the average of the scores awarded by the three raters.

(4) *Reliability of the scales*

The reliability of each rating scale was found by intercorrelating the scores of the three sets of judges and taking the average of the three coefficients obtained. This gave:

Block-in scale (A):	'*P*'	$r = 0{\cdot}82$
	'*w*'	$r = 0{\cdot}76$
Hypothetical situation scale		$r = 0{\cdot}52$
Marston scale		$r = 0{\cdot}69$

(The reliability of the Marston scale itself was found by correlating the sum of the scores on odd traits with the sum on even traits. The reliability of the half-test was then 0·83, or for the whole scale, 0·91.)

The reliability of the rating scales is unfortunately very much lower on the whole than that of the test results. In part this may be due to the usual discrepancies of ratings of personal characteristics; it may also be due to the small number of subjects and to the fact that each subject had her own judges. Had the subjects been more in number, and had they been rated by the same three judges, it is possible that the reliability would have been higher.

(5) *Intercorrelation of rating scale results*

From all the rating scales used, eight measures were finally available —four of 'persistence', one of '*w*', one of introversion and two self-

ratings. The reliability of the self-ratings was found to be lower throughout than that of ratings of judges, and so these are not described in the present report.

In Table XIX the correlations given in brackets are only approximate, being the coefficients over and above those to be expected from chance

Table XIX. *Rating scale intercorrelations*

Rating scale		1	2	3	4	5	6	Average
'*P*': hypothetical situation scale	(1)		0·72	0·60	0·56	0·74	0·58	0·64
'*P*': block-in scale	(2)			(0·40)	0·38	0·46	0·47	0·49
'*w*': block-in scale	(3)				0·35	0·50	0·28	0·43
Introversion score	(4)					(0·43)	(0·49)	0·44
Question 5 of Marston scale*	(5)						0·71	0·57
Question 16 of Marston scale†	(6)							0·51

Average $r = 0{\cdot}51$.

* Question 5 is: "Tends towards activities requiring care; good in detail."

† Question 16 is: "Does not abandon an activity readily."

alone, due to the overlapping of questions. The coefficients are now all positive, with an average correlation of 0·51. The measures are not of course entirely independent of each other, although overlapping has been ruled out as far as possible. A good degree of correspondence between 'persistence' as described abstractly and as described in specific situations was expected and hoped for; and the coefficient of 0·72 is satisfactorily high in view of the arbitrary selection of the 'hypothetical situations' and the unexplained trait descriptions of the block-in scale. The best-correlating measure is the hypothetical situation scale (average $r = 0{\cdot}64$). The lowest is the '*w*' measure, which is perhaps not surprising in view of the abstract nature of the '*w*' descriptions.

V. Correlation of rating scales and test results

(1) *Ratings and variability scores*

The previous sections have been concerned with character ratings on the one hand—their internal consistency and intercorrelation—and with test results on the other. The main aim of the experiment was, however, to bring the two together and see how, if at all, these character traits would be revealed in accuracy, speed or variability of work.

Table XX gives the correlation of all the ratings (neglecting only the self-ratings) with the variability scores, i.e. with low variability. Considering the arm-movement and ball-and-slot tests together, it is clear

Table XX. *Variability and persistence ratings*

Rating scale	Test I (arm-movement)			Test II (ball-and-slot)			Test III (character judgement)		
	A	B	Av.	A	B	Av.	A	B	Av.
'*P*': hypothetical situation scale	−0·14	−0·04	−0·09	0·22	0·08	0·15	0·30	0·39	0·34
'*w*': block-in scale	0·07	−0·18	−0·05	−0·20	−0·30	−0·25	0·33	0·23	0·28
'*P*': block-in scale	0·25	−0·05	0·10	−0·06	−0·18	−0·12	0·24	0·19	0·21
Question 5 (Marston)	0·05	0·16	0·10	0·20	0·08	0·14	0·47	0·50	0·48
Question 16 (Marston)	0·26	0·08	0·17	0·07	−0·02	0·02	0·21	0·32	0·26
Average	0·10	−0·01	0·04	0·05	−0·07	−0·01	0·31	0·33	0·32
Introversion	−0·21	−0·18	−0·19	−0·25	−0·39	−0·32	0·17	0·16	0·16

that there is no significant relationship present; no one of the five persistence measures shows any consistent correlation with variability of work, although the negative coefficients with introversion suggest that possibly some or all of the extravert qualities tend to produce successful motor performances such as are required in these tests, and so indirectly produce low variation.

This lack of correlation was to be expected in two tests where the amount of error correlated highly with variability. It seems that the relative liability to error so masked comparative variability that the average deviation from the mean was not a true variability score, i.e. it was not an independent score, but was dependent upon the error scores. The highly inaccurate subjects, probably because of their knowledge of success and failure, had to vary their judgements more than those, more accurate, who were more satisfied with their results.

Therefore an attempt was made to unmask true variability scores to a small degree at any rate, by dividing the variability score of each subject by her error score. This was done in both the arm-movement and ball-and-slot tests. In the arm-movement test, however, the extremely inaccurate subjects then obtained the lowest relative variability scores, simply because of the size of the denominator in the ratio variability/error. The relative variability score was therefore taken as

$$\frac{\text{rank order on variability scores}}{\text{rank order on error scores}}.$$

The correlations with the rating scales were then repeated for relative variability and were as shown in Table XXI. In this table the coefficients for the continuous work of the arm-movement test are all negative, but it must be remembered that the A period of this test was

Table XXI. *Variability relative to error*

Rating scale	Test I		Test II	
	A	B	A	B
'*P*': hypothetical situation scale	−0·25	0·33	0·65	0·37
'*w*': block-in scale	−0·17	0·21	0·43	0·39
'*P*': block-in scale	−0·06	0·50	0·37	0·21
Question 5 (Marston)	−0·19	0·17	0·43	0·23
Question 16 (Marston)	0·01	0·14	0·33	0·15
Average	−0·13	0·27	0·44	0·27
Introversion	0·01	−0·10	0·32	0·47

found to be unsatisfactory—its correlation with the B period was low and it correlated poorly with variability on the other tests.

The other coefficients now show positive correlations with the rating scale results. Their actual magnitude is of little value because of the small number of subjects, but the consistently positive correlation does show that there is a tendency for lack of variability to be associated with very varying measures of 'persistence' and '*w*'—measures whose intercorrelation averages only 0·51.

It is of some interest to compare the values of the rating scales from the point of view of their correlation with the test results. The average correlation on A and B periods on all three tests has therefore been taken and is as shown in Table XXII. The best correlation is with persistence

Table XXII

Rating scale	Periods		Average
	A	B	
'*P*': hypothetical situation scale	0·47	0·36	0·41
'*w*': block-in scale	0·38	0·26	0·32
'*P*': block-in scale	0·30	0·30	0·30
Question 5 (Marston)	0·45	0·30	0·37
Question 16 (Marston)	0·27	0·20	0·23
Average	0·37	0·28	

as exemplified in specific, imaginary situations. This is also the rating scale that has the best average intercorrelation with the other scales (0·64), although its internal consistency is not good (0·52). Its correlation with the test variability scores, except in the unreliable I A period, never drops below 0·29, and with the continuous work periods of the character judgement test is as high as 0·65. This suggests that the most fruitful form of rating scale to take would be a 'specific situation' scale. It would, however, be constructed in more detail than the hypothetical situation scale, and on the basis of the best intercorrelating questions.

If this were done, the internal consistency might be greatly improved, and the correlation with variability might also be higher.

The above results show that there seems to be no greater correspondence after a long time of testing than in the first half an hour's work. This is not true of the arm-movement test, in which the first half-hour seems to involve the interplay of many factors, but the ball-and-slot and character judgement tests correlate with variability equally well in the A periods of 30 min. continuous work as in the later B periods of 3 min. work a day for 10 days. It cannot be said that this would be true for other kinds of work, and it would be very hard to frame any law on the length of testing time necessary in judgement tests generally; but from the point of view of further work it would be possible to design tests such as these latter two, which need not be taken to any great length of testing time for them to give reliable variability scores.

(2) *Ratings and error scores*

The correlations of rating scales and actual error scores were obtained but found to be entirely negligible (average $r = -0{\cdot}095$). The table of correlations is therefore not given.

(3) *Ratings and speed*

The average correlation of persistence ratings with slowness was $-0{\cdot}055$, and here again the table of correlations has been omitted. Contrary to popular belief, the careful, persistent workers (i.e. those rated as such) did not prove to be any slower than those rated as more impulsive and careless.

In the intercorrelations between measures of speed on the three tests, no evidence was found for a speed factor, speed being consistent throughout a test but not overlapping with the speed scores on any other test. Here speed seems to be independent of the wide character qualities of the rating scales; so that, from a diagnostic point of view, the speed with which a subject works is no indication of what his speed will be on any other kind of work, nor is it any indication of character.

(4) *Correlation of variability with the whole block-in scale*

Variability has already been correlated with traits of the block-in scale which were pooled to form 'P' and 'w' estimates. The full correlation of the character judgement test (III) with all the thirty traits of the scale was also obtained, this test being chosen as that yielding variability scores uncomplicated by error. This full list includes the 'self' traits and

others which have so far been left out of account. Of the thirty correlations obtained the best-correlating traits, those above 0·20, are shown in Table XXIII.

Table XXIII. *Correlation of traits with variability*

Positive correlations		Negative correlations	
Conventionality	0·62	Originality	−0·33
Strict adhesion to principles	0·29	Speed of work	−0·30
Pugnacity	0·25	Talkativeness	−0·29
Thoroughness in routine work	0·23	Imagination	−0·27
Strength of will	0·22	Quickness of understanding	−0·27
		Shyness	−0·27
Average $r = 0{\cdot}32$.		Average $r = -0{\cdot}29$.	

There are several seeming discrepancies here. In the positive correlations 'pugnacity' (in Webb's research 'readiness to become angry') is not generally associated with the '*w*' qualities, and in the negative correlations, 'shyness' and 'talkativeness' are equally correlated. But, apart from these instances, the positive correlations are on the whole with traits of a '*w*' type. The four best-correlating traits are in fact the very ones selected, quite independently and before this full correlation with the block-in scale was done, to represent '*w*'. The traits correlating negatively, on the other hand, are also easily recognizable as non-'*w*' traits—in many respects the reverse of '*w*'—'originality' and 'imagination' in place of 'conventionality'; 'quickness' and 'speed' instead of 'thoroughness'.

Thus there is evidence from several sources that variability of judgement can be measured and found to be consistent over a long-continued time of testing. There is evidence for a group factor of variability present even when the influence of accuracy has been eliminated; and last, and most important, there is evidence for a consistent, low correlation of variability with measures of 'persistence' and '*w*'. There is no information as to the closeness of relationship; the evidence shows only its existence. Such evidence rests, not upon the size of the coefficients, which would be of no value in results from a group of only eighteen subjects, but upon the general trend of the coefficients.

VI. Summary of main results

1. The relations of speed, error and variability to each other were not general, but dependent upon specific conditions in each test.

2. Speed within each test was highly consistent throughout half an hour's work and the subsequent ten 3 min. daily periods, but there was no evidence for a speed factor underlying the scores on the three tests.

3. There was evidence for a factor of variability, present in all three tests, but in much greater evidence in the later 3 min. a day testing than in the earlier continuous half-hour's work.

4. The rating scales of 'persistence' and '*w*', although less reliable than the test results (as judged by the reliability coefficients) showed consistently positive correlation.

5. Low variability on the character judgement test correlated positively with all the 'persistence' and '*w*' measures at four points in the progress of testing, giving twenty coefficients in all. The small number of subjects makes the exact values of the coefficients of little value, but the general trend of positive correlation seems to show that 'steadiness' of character is associated with low variability.

6. Low variability, taken relative to the actual error scores, correlated positively with '*w*' and 'persistence' estimates in the latter half of the arm-movement test and in both types of work in the ball-and-slot test. These estimates did not, however, correlate with the crude variability scores on these tests.

7. Low variability correlated positively with introversion, although the coefficients are not high enough to be individually significant, nor sufficient in number to show the general trend of correlation.

8. The hypothetical situation scale had the best average correlation with the variability scores on all three tests. It seems therefore that the best type of rating scale is one of specific, imaginary situations rather than one listing abstract trait-names.

9. The relationship of accuracy to the 'persistence' scales was entirely negligible, though there is a slight positive correlation with extraversion.

10. There was similarly no significant relationship or trend of correlation between speed and 'persistence'.

REFERENCES

(1) Webb, E. (1912). "Character and intelligence." *Brit. J. Psychol. Mon. Supp.* I, No. 3.

(2) Cox, J. W. (1934). *Manual Skill.* Cambridge University Press.

(3) Walton, R. D. (1936). "Relations between the amplitude of oscillations in short period efficiency and steadiness of character." This *Journal,* XXVII, 181–8.

(4) Hollingworth, H. L. (1925). "Correlations of achievement within an individual." *J. Exp. Psychol.* VIII, 190–208.

(5) Asch, S. E. (1932). "An experimental study of variability in learning." *Arch. Psychol.* No. 143, 1–54.

(6) MAY, M. A. & HARTSHORNE, H. (1929). *Studies in Service and Self-control.* New York: Macmillan Press.
(7) CRUTCHER, R. (1934). "An experimental study of persistence." *J. Appl. Psychol.* XVIII, 409–17.
(8) HOWELLS, T. H. (1933). "An experimental study of persistence." *J. Abn. Soc. Psychol.* XXVIII, 14–29.
(9) GOBAT, J. D. (1937). "Error and variability of judgement in routine work; their relation to assessments of character." Thesis for Ph.D. degree. Deposited British Museum.
(10) GUILFORD, J. P. & HUNT, J. M. (1931). "Some further experimental tests of McDougall's theory of introversion-extraversion." *J. Abn. Soc. Psychol.* XXVI, 324–32.
(11) VERNON, P. E. (1933). *Studies in Expressive Movement.* London: Macmillan.
(12) GARNETT, J. C. M. (1919). "General ability, cleverness and purpose." This *Journal,* IX, 345–66.
(13) BROWN, W. & THOMSON, G. H. (1925). *The Essentials of Mental Measurement.* Cambridge University Press.

(*Manuscript received* 10 *February* 1938)

SOME RELATIONS BETWEEN REPRODUCING AND RECOGNIZING PROSE MATERIAL

By O. L. ZANGWILL
Cambridge Psychological Laboratory

I. *Introduction* (pp. 370–371).
II. *Experimental method* (pp. 371–376).
(1) *Material and instructions* (pp. 371–372).
(2) *Conditions of reproduction* (p. 372).
(3) *Material for recognition* (pp. 372–375).
(4) *Instructions for recognition* (pp. 375–376).
(5) *Subjects* (p. 376).
III. *Principal results* (pp. 376–378).
IV. *Some qualitative observations* (pp. 378–380).
(1) *Construction and style in relation to recognition* (p. 378).
(2) *Familiarity and unfamiliarity* (p. 378).
(3) *The dominance of details* (pp. 379–380).
V. *Positive and negative relations in recognition and recall* (pp. 380–381).
VI. *Summary and conclusions* (pp. 381–382).
References (p. 382).

I. Introduction

In an earlier paper(3), some evidence was produced to the effect that the various varieties of constructive and dynamic processes which have been shown to operate in remembering are not without influence on recognition. This evidence was neither certain nor conclusive; yet the relations which we were able to establish between some aspects of recognition and recall seemed of sufficient interest to justify further study.[1]

The present investigation departs from its predecessor in two important respects. In the first place, connected prose passages were used instead of simple figures; in the second place, the conditions were arranged with a view to minimizing the effects upon recognition of interpolated recall. The technique of repeated reproduction was entirely discarded, and the conditions were, in consequence, better adapted to the study of anomalies in the process of recognition itself.

[1] It is perhaps worth remarking that Hanawalt's extensive results(2) have largely confirmed the modest observations recorded in our preliminary study.

The general plan of the experiment was identical with that of the earlier investigation. A number of untrained subjects were instructed to read through an unfamiliar prose passage and to make no attempt to memorize the material. Three months later, a reproduction of this passage was requested from each of the subjects. Unknown to the latter, the material of these reproductions was utilized by the experimenter in the preparation of special passages for recognition. Some time later, any subject who had furnished a reasonably detailed recall was provided with three superficially similar prose passages. These comprised (*a*) the original narrative; (*b*) a passage of approximately equal length and not altogether unlike the former in style and content; and (*c*) a narrative constructed by the experimenter in such a way as to include the most salient inaccuracies of the individual's own reproduction. The various transformations and inventions which had originated in recall were welded together into the form of a coherent narrative bearing a strong resemblance to the original in style and construction. Each subject was instructed to read through the three passages, and to indicate in a prescribed manner the one which he believed to be the original. He was also asked to specify the certainty of his judgement and the factors upon which he believed it to be based.

II. Experimental method

(1) *Material and instructions*

The passage selected was a portion of a Hindu legend reproduced by Oesterreich in his book on *Possession*. It was chosen as a suitable vehicle for the general tendencies of 'conventionalization' whose operation in recall has been so strikingly illustrated in the studies of Bartlett. The narrative reads as follows:

There is a town of the name of Vatsamân, where lived a Brahman called Kecava who was wise, but poor. His wife, who was called Karagarâ (i.e. poison-giver), behaved so ill towards everyone that even a demon, who lived on a tree in the house, fled into the desert for fear of her. Meanwhile the Brahman was no longer able to bear the wickedness of his wife and went away also. On the way to the desert the demon saw him and said to him: "I wish to-day to offer thee hospitality." When the Brahman heard this he was afraid. "Fear not," said the demon, "for I used to live upon a tree in thy house, but I fled to this spot for fear of Karagarâ, and since thou hast long had to do with me as landlord I will do good to thee. Go thy way to the village of Mrigavatî (that is to say, rich in Gazelles); I will take possession of the daughter of the city, Mrigalotschanâ (signifying gazelle-eyed), and will not let myself be driven out by any exorcist; thou alone, when thou comest, shall drive me out with a look." Having spoken thus the demon entered into the royal virgin. Meanwhile the Brahman went

to the royal city Mrigavatî and having heard the herald he went to the royal palace, but in vain he did all that magicians are accustomed to do and uttered his conjurations, the demon did not leave the maiden. When the Brahman saw that in no other way would the demon come out, he cried: "In the name of Karagarâ, come forth!" The demon replied: "See, I am coming forthwith!" and he immediately came out. Then the King gave to the Brahman the half of his kingdom and his daughter in marriage.

Each subject received a copy of the above story together with the following instructions:

Please read through the enclosed story twice, at your normal rate of reading. As soon as you have thus read it, please destroy it *without fail.* You will not be asked to reproduce the story at any stage of the experiment, *so please make no attempt to learn the whole or parts of it by heart.*

Further instructions will be forwarded to you in due course.

The entire experiment was carried out without direct contact between the experimenter and the subjects, all communication being reciprocally postal. Thus although the requisite control of the conditions depended entirely upon the goodwill and co-operation of the subjects, the somewhat artificial atmosphere of the usual laboratory experiment was largely avoided.

(2) *Conditions of reproduction*

After a lapse of three months, the following communication was sent to the subjects:

Some time ago, you received a short story which you were asked to read and then to destroy. We should be grateful if you could write down on the enclosed paper as much of this story as you can remember and forward it to us in the enclosed envelope. If you are unable to remember the story coherently, please write down any detail or indeed anything connected with the story that you may be able to recall. Should you be unable to remember anything whatsoever, please return the enclosed paper stating this fact.

We should like to apologize for our apparent infraction of the original instructions, but it was essential for our purposes that intentions to remember should be rigidly excluded.

The instructions for the final stage of this experiment will be forwarded to you very shortly.

(3) *Material for recognition*

The mode of construction of the material provided for recognition has already been mentioned. It will be remembered that each subject received three passages, labelled A, B and C respectively. Version C was the original story; version B a specially prepared narrative based on the reproduction; and version A an unfamiliar narrative introduced as a kind

of control. For purposes of comparison, the narrative A is reproduced below:

There is a town of the name of Vatsamân where lived a Brahman called Kecava who was wise but poor. The good Brahman and his family were greatly tormented by a demon, who lived upon a tree in the house. Earnestly the Brahman pleaded with the demon, and finally the latter agreed to obey his bidding and departed from this place. When the demon had come forth from the tree, he went into the town of Karnavatî (that is to say, the town with ears) and took possession of the queen, who was half-sister on the father's side of the King Madana and who was called Sulotschanâ, (signifying lovely-eyed). Greatly tormented by the demon, the queen became like a skeleton. Then the king, whose name was Satrughna (slayer of enemies), sent to the King Madana and begged him to send the wise Brahman Kecava, who had rid the town Vatsamân of a demon who lived in a tree. At the request of Madana, Kecava came to Karnavatî to the possessed queen. But when the demon perceived him he insulted and threatened him: "It is enough that I have done thy bidding once; now take care and look to thyself!" When the Brahman heard this, he recognised that it was the same demon. When the Brahman saw that in no other way would the demon come forth, he uttered his magical conjurations. The demon was seized with fright and immediately left the queen. The Brahman, covered with honours by the King, made his way back to Mrigavatî.

Seven of the nineteen subjects gave recalls sufficiently detailed and coherent to furnish the basis of special recognition versions. One of the seven indeed provided sufficient material for the construction of two such versions (in this case labelled B and D respectively).

As illustrations of the method used in construction of the recognition material, two complete examples may be given:

Subject K: Recall version. I remember the scene of the story as being set in the Kashmere-Punjab district in a village set in well-wooded foothills. Its name began with M. There was a middle-aged man of rustic profession (e.g. woodcutting?) therein, with a horrid mother-in-law (or step-mother). Both their names began with M. From the forests near by a were-tiger (or maybe just an ordinary tiger) made depredations on the villagers and their gardens. The wood-cutter somehow killed it. At this point the step-mother seems to go out of the picture. Perhaps she was identified with the tiger. The next scene is the charcoal burner (*sic*) leaving for some reason (the old mother's malign influence?) and going south into the environs of an important city of the Plains, where he met a princess, whom however I cannot remember him actually marrying, though that seemed to be the natural outcome if the story had not ended before it could happen; her name began with an M.

Recognition (B) version. There is a village of the name of Vatsamân which is surrounded by hills and woodland. Here lived Kecava, who though but a poor wood-cutter, was a Brahman and very wise. Kecava and his wife suffered greatly under Karagarâ (i.e. poison-giver), mother-in-law to Kecava, who behaved so ill towards everyone that even the leaves of a tree in the courtyard withered and no flowers would bloom. At that time, the villagers were suffering greatly from the

depredations of a tiger who stole forth at nightfall from the forests, and whose ravages struck terror into the hearts of the humble villagers. One evening, the Chief of the village Elders approached Kecava and said to him: "Never before has mortal beast so ravaged our peaceable villagers. Our Elders have held council, and believe this to be the work of the evil woman Karagarâ, the Giver of Poison. It is she who takes possession of the tiger and leads it to ravage our village." Kecava went forth, and after many weeks of search, he encountered the tiger, that had ravaged the village of Vatsamân. A great struggle took place, and the tiger was slain by the hand of the Brahman. The wanderings of Kecava had carried him south into the plains and soon he reached the Royal City of Mrigavatî. The people of this city were in mourning because the Royal Virgin, Mrigalotschanâ (which signifies gazelle-eyed) had been possessed by a demon. Having heard the herald, Kecava proceeded to the Royal Palace, but in vain did he do all that magicians are accustomed to do; the demon did not leave the maiden. When the Brahman saw that in no other way would the demon come forth, he cried: "In the name of Karagarâ, come forth!" Immediately the demon came out and the maiden became once more her true self. Then the king gave to the Brahman the half of his kingdom and his daughter in marriage.

Subject O: Recall version. A Brahmin was having supper one night with his wife when he heard a loud wailing and screaming in the garden. Rushing out, they discovered a demon caught in the branches of a pomegranate tree from which he was unable to extricate himself. The woman was terrified and rushed into the house again, but the Brahmin stood his ground and held conversation with the demon. In the course of this the spirit promised that if he were released from the tree so that he could return to his own works, he would be at any time at the order of the Brahmin. At this the Brahmin fetched a ladder, cut away with an axe the encircling branches, and the demon flew away.

Shortly after this the Brahmin's wife died and he departed from the place in which he lived. He walked for days through a desert until at last he came to a city which was called "The Pools of the Gazelles", and which was ruled by a King the name of whose daughter, the Princess, translated means "Gazelle-eyed". The people of that city were in mourning because the princess was possessed by a demon, and although two other famous sorcerers were there, their efforts to dislodge the demon were in vain. The Brahmin approached the princess and said in a loud voice "Come forth!" "I come", replied a voice which was the voice of the spirit which had been caught in the pomegranate tree. The princess was immediately perfectly normal.

Recognition (B) version. There is a town of the name of Vatsamân, where lived a Brahman called Kecava who was wise but poor. His wife, who was called Karagarâ (i.e. the poison-giver), behaved so ill towards everyone that even a demon who lived on a pomegranate tree in the garden, flew away wailing and screaming into the desert for fear of her and returned to his own works. Meanwhile the Brahman was no longer able to bear the wickedness of his wife and went away also. The demon met him and said to him: "I wish to offer thee hospitality." When the Brahman heard this he was afraid. "Fear not," said the demon, "for I used to live upon a pomegranate tree in thy house, but I fled to this spot for fear of Karagarâ. And since thou hast long had to do with me as my landlord, I will do good to thee. "Go thy way to the city Mrigavatî (that is to say, 'The Pools of the Gazelles'); I will take possession of the daughter of the city, Mrigalotschanâ (signifying Gazelle-eyed), and will not let myself

be driven out by any exorcist. Thou alone shalt drive me out with a look when thou comest." Having spoken thus, the demon entered the Royal Virgin. Meanwhile the Brahman walked for many days through a desert until at last he came to a city which was indeed Mrigavatî. The people of this city were in mourning, because the princess was possessed of a demon, and although two famous sorcerers were there, their efforts to dislodge the demon were in vain. Then the Brahman approached the princess and cried in a loud voice: "In the name of Karagarâ, come forth!" "I come!" replied a voice which was the voice of the spirit who had lived in the pomegranate tree. The princess was immediately perfectly normal. Then the King gave to the Brahman the half of his kingdom and his daughter in marriage.

Recognition (D) version. There is a town of the name of Vatsamân, where lived a Brahman called Kecava who was wise but poor. As Kecava and his wife, called Karagarâ, were sitting one night at their supper, they heard a terrible wailing and screaming. Rushing out from the house, they discovered a demon caught among the spreading branches of a large tree which grew in the courtyard. Karagarâ was overcome with terror, and fled back trembling into the house, but the wise Brahman approached the demon and held conversation with him. "If thou releasest me from the imprisoning branches of this tree," pleaded the demon, "then will I promise to serve thee faithfully and to be ready at thy service whenever thou shouldst need me." The Brahman then cut away the encircling branches and the demon flew away. Shortly after this, the wife of Kecava perished, and the good Brahman departed from the place Vatsamân in which he had lived so long. On his way to the desert, he passed through the royal city of Mrigavatî (that is to say, rich in gazelles). Here he found the Royal Virgin, Mrigalotschanâ (signifying gazelle-eyed), to be possessed by a demon. Having heard the herald, he went to the royal palace, but in vain he did all that magicians are accustomed to do and uttered his conjurations; the demon did not leave the maiden. When the Brahman saw that in no other way would the demon come out, he cried in a great voice: "Come forth!" "See, I am coming forthwith", replied a voice which was the voice of the demon that had been imprisoned in the Brahman's tree. Immediately he came forth, and the Royal Virgin became once more herself.

The above examples furnish excellent illustrations of constructive recall. In the case of Subject O, it will be seen that the more significant deviations from the original narrative (18 in all) are included in version D, whereas the relatively minor inaccuracies of detail have been introduced into version B.

(4) *Instructions for recognition*

Between two and three weeks after the recall versions had been submitted, each of the aforementioned seven subjects received the following instructions:

Enclosed are a number of stories, *one* of which you were asked to read as the initial stage of the experiment. You are requested to attempt to recognize (i.e. to identify) the original story. Each story is marked A, B or C. Please place a cross on the enclosed card against the letter corresponding to that of the story which you believe to be the

original. Please indicate your degree of certainty by writing 'certain', 'probably', 'possibly' (as the case may be) alongside the cross. Any introspective reports about the processes that culminated in your verdict will be welcome.

(5) *Subjects*

The subjects comprised nineteen highly educated adults of both sexes and of diverse age and occupation. Of the seven subjects who received recognition versions, five were women and two men. None of the subjects had at any time followed courses in experimental psychology and all but one were unfamiliar with Bartlett's work on remembering.

III. Principal results

The general outcome of our inquiry is best presented in the following manner. In the present section, certain outstanding results will be given in tabular form. In the following section, a number of qualitative observations relating to the general conditions of recognition will be discussed. Finally, in the penultimate section, we shall attempt a comparison of our results with those of the earlier investigation.

The principal results are as follows:

(1) The original story (recognition version C) was identified with unqualified conviction by only one of the seven subjects to whom recognition material was submitted.

(2) Two subjects identified the original story with a fairly high degree of certainty. Both of these nevertheless admitted that version B (that based upon the respective reproduction) might just possibly be the original.

(3) Three subjects affirmed that none of the three versions was the original. Two of these three actually expressed the belief that the original story had included passages from both the B and the C versions.

> I imagine the stories B and C both contain parts of the original one. *Subject H.*
>
> Finally I said the story I saw was C with the first third of B worked into it but here left out. In other words, none of these three is the exact story. *Subject K.*

The third subject left her card blank, but her introspective report furnishes an admirable example of the reconstructive character of recognition under these conditions. It may well be quoted intact:

> None of these stories is the original story which I saw, although D is the nearest version of it. Yet the lame ending of D proves that that version has been tampered with. A has a different characterization from the original story, and I most certainly never saw it before. B has the pomegranate tree which I remember and the correct translation of the city's name, but the original story was not as long as B. As a matter

of fact, when I look at these again, I begin to believe that the original version was C, and that I chose D because it was likest to my own version. *Subject O.*

Another subject, though ultimately deciding in favour of C, wrote the following note in her report:

If I didn't believe in the integrity of the compiler of this test, I should think that none of these three is the original. As it is, the nearest is C.

(4) All subjects were convinced (though on very different grounds) that the control story (A) did not constitute the original.

(5) It is interesting to consider the nature of the verdict in relation to the extent of transformation in recall. It must however be clearly understood that the 'points of deviation' listed in Table I are not to be treated as psychologically comparable. In some cases, a recorded 'deviation' consists only in some transformed, or imported, detail; but in other cases, a single 'deviation' may automatically involve widespread change in plot and sequence. Deviations of the latter type had of course to be taken into account in constructing the recognition material and they frequently gave rise to very striking distortions of the original narrative.

Table I

Subject	Points of deviation in recall introduced into recognition material	Nature of the verdict
D	2	C, probably; B, just possibly
M	3	C, certain
R	3	C, almost certain; B, just possibly
H	6	Verdict initially in favour of C, but original finally believed to embody parts both of B and C
V	9	Verdict ultimately in favour of C, but tendency to treat original as embodying parts both of B and C
K	9	No final positive verdict on any story. Original treated as a blend of B and C
O	15 (B) 18 (D)	No final positive verdict on any story. Considerable uncertainty and confusion in regard to versions B, C and D

Table I indicates clearly that the nature of the verdict varies significantly with the number of deviations manifested in recall and introduced by the experimenter into the recognition material. We are able in fact to conclude that where reproduction is highly constructive, recognition of the original passage under our experimental conditions is likely to

become both difficult and uncertain. Moreover, there is evidence of a strong tendency in some of our subjects to treat the original as a blend. They affirm that certain passages in the B versions were present in the original beyond all reasonable doubt. This naturally gives rise to the suspicion that none of the three versions is the original.

IV. Some qualitative observations

(1) *Construction and style in relation to recognition*

Bartlett ((1), p. 81) has pointed out that the presentation of complex verbal material tends to evoke immediate and characteristic responses to general construction, syntax, rhythm and style. These factors are, he finds, extremely readily responded to, but extremely rarely reproduced with any fidelity. Bartlett's contention was amply borne out in the present study. In particular, it was observed that extremely few of the subjects made any serious attempt to recapture the style of the original narrative in their reproductions. But it is important to note that an impression of style may nevertheless persist, and be of significance in directing recognition. This is well shown in the following extract from the report of subject D:

> Much troubled by the dog in B, but C seems to have the shape and style of the original rather more. (The dog mentioned in the report was an error in recall inserted by the experimenter into version B.)

In general, it seems that responses to style and construction, while of little importance for reproduction, are not without significance in relation to recognition. This conclusion is certainly in accord with the experience of the skilled literary critic.

(2) *Familiarity and unfamiliarity*

Feelings of familiarity or unfamiliarity were seldom mentioned by our subjects in their reports on recognition. In a few cases, isolated details, or phrases in the dialogue, were said to call out immediate reactions of familiarity. But in no single case was the narrative as a whole said to give rise to an experience of this character. In general, our results lend little support to the view that recognition invariably depends upon the appearance in consciousness of an 'ultimate and irreducible' feeling of familiarity.

(3) *The dominance of details*

Bartlett ((1), p. 31) has pointed out that all perceiving involves a kind of unwitting analysis. Subjective and objective factors together combine to produce a diversification of emphasis in the perceptual field. There are always some features of the perceptual situation which take a lead over the others. There are always some details likely to stand out and influence what is perceived more than the rest. This distribution of emphasis within a perceptual field is the phenomenon of *dominance*.

The significance of dominance is by no means restricted to perception in the narrow sense. The phenomenon plays an important part in the shaping of responses remote in time from the original perceptual activity. "These dominant details", writes Bartlett, "are a kind of nucleus about which the rest cluster. They set the stage for remembering."

Our present investigation provides some data bearing upon the phenomenon of dominance and its implications for recognition. In the first place, it was found that outstanding detail appears to provide the principal clue in identifying prose material under our experimental conditions. Indeed the outcome of a judgement upon any given passage seems to be almost entirely a function of these responses to detail. The recognition of a striking detail appears to carry with it an overwhelming conviction that the entire passage must be familiar.

Some examples of the dependence of the verdict upon responses to detail may now be given:

(1) The presence in version C of the term "poison-giver" and of the phrase "the demon entered into the royal virgin" led Subject H to an immediate positive verdict upon this passage. Her verdict was subsequently qualified only through the inclusion in version B of a detail, which though absent in the original, nevertheless appeared to acquire a certain relative dominance. This is discussed below.

(2) The final selection of version C by Subject V was principally due to a phrase in the dialogue which occurs in this version alone.

A second finding of some interest is that detail introduced by a subject in remembering may apparently function as dominant and significantly influence the process of recognition. Indeed the acceptance of an invented detail as genuine was the most frequent cause of error in discrimination. This is illustrated by the following extracts from individual reports:

(1) Version B has the pomegranate tree which I remember and the correct translation of the city's name. *Subject O.*

At no point in the original was there any mention of a "pomegranate tree" and the alleged "correct translation" of the name of the city itself originated as an inaccuracy in recall.

(2) A point that also remained in my mind strongly in the original was the description of "the leaves of the tree in the courtyard withering and no flowers blooming". Well, that sentence comes in B, so I imagine the stories B and C both contain parts of the original one. *Subject H.*

The passage in question does not occur in the original narrative.

(3) I remember definitely the words: "In the name of Karagarâ, come forth!" On the other hand, I *did* think he was a woodcutter. If I didn't believe in the integrity of the compiler of this test, I should think that none of these three is the original. *Subject V.*

The substitution of 'woodcutter' for 'Brahman' characterized the reproduction of this subject.

(4) On the negative assumption that I should have remembered one or two phrases in *B*, e.g. "The voice was the voice of the traveller", I decided it was C. *Subject R.*

The last extract furnishes a good example of the negative implications of this principle of dominance. The subject in question is obviously aware of the existence of a selective process in perception and recall.

The above illustrations enable us to conclude that detail introduced into the narrative in recall may operate as dominant in much the same manner as outstanding material genuinely present in the original. This conclusion might have an important bearing upon questions of testimony and paramnesia in daily life.

V. Positive and negative relations in recognition and recall

In our earlier investigation, positive relations[1] were found between persistent deviations in a given series of reproductions and the corresponding characteristics of a figure recognized as the original. Positive relationships are not universal, but were observed in 62% of the total classified errors. It was further remarked that both positive and negative relationships might sometimes be demonstrated within the same objective figure.

If we take into account the vast differences in the nature of the material, in the conditions of presentation and reproduction, and in the duration of the intervals between the various stages of the experiment, it must be admitted that the results of the present research are in

[1] The expressions 'positive' and 'negative' relation have been defined elsewhere by the writer ((3), p. 266). They have been adopted, with some modification, by Hanawalt ((2), p. 68).

remarkable agreement with those of the earlier investigation. In both cases, deviations originating in recall are treated as genuine characteristics of the original and are recognised as such with a high degree of certainty. In other words, positive relationships have now been demonstrated not only in the case of relatively meaningless diagrams, but also in the case of complex and meaningful prose material. It has also been shown that some of the deviations are not treated as characteristics of the original narrative. These are examples of negative relationships very similar in kind to those reported in the earlier investigation. Lastly, it has been shown that the existence both of positive and negative relations within a given passage gives rise to considerable difficulty in expressing a judgement of recognition upon the passage as a whole. Precisely similar difficulties were encountered by the subjects in the earlier study.

VI. Summary and conclusions

The investigation reported in this paper represents an attempt to study some interrelationships between processes of reproducing and recognizing prose material. The principal conclusions are as follows:

(1) Recognition of a prose passage may be rendered exceedingly difficult by submitting it together with specially prepared material based upon a reproduction of the original previously furnished by the subject. In some cases, recognition seems to be virtually impossible; in others, there is a marked tendency to conclude that the original passage must have included a number of the fictitious points which originated in recall. It is suggested that these observations may have some bearing upon paramnesia in daily life.

(2) Factors of style and construction, though rarely reproduced with any accuracy, are nevertheless influential as cues in selective recognition.

(3) Recognition is frequently based almost entirely upon the response to some outstanding detail. Errors in recognition for the most part derive from similar responses to importations and constructions. The significance for recognition of 'dominant' detail is thus fairly considerable.

(4) Distinct feelings of familiarity are seldom reported. Where they occur, they are said to be aroused by relatively isolated details and not by the original passage as an undivided whole.

(5) It is pointed out that the results of this investigation parallel very closely those of an earlier study in which simple designs were used as the

experimental material. Both positive and negative relationships are found to exist between the recognition and the recall of narrative prose.

I should like to express my thanks to Mr R. N. Sinha for some help in the selection and preparation of the material.

REFERENCES

(1) Bartlett, F. C. (1932). *Remembering.* Cambridge University Press.
(2) Hanawalt, N. G. (1937). "Memory trace for figures in recall and recognition." *Arch. Psychol., N.Y.*, no. 216.
(3) Zangwill, O. L. (1937). "An investigation of the relationship between the processes of reproducing and recognising simple figures, with special reference to Koffka's trace theory." *Brit. J. Psychol.* xxvii, 250–76.

(*Manuscript received* 6 *June* 1936)

THE EFFECTS OF ERRORS OF MEASUREMENT ON CORRELATION COEFFICIENTS

By ROBERT H. THOULESS

I. *The correction for attenuation* (pp. 383–387).
II. *Objections to correction for attenuation and difficulties in its use* (pp. 387–391).
III. *Conditions in which coefficients of correlation should and should not be corrected for attenuation* (pp. 391–396).
IV. *Partial correlation* (pp. 396–400).
V. *Summary* (pp. 400–401).
Appendix (pp. 401–402).
References (p. 403).

I. The correction for attenuation

It was first pointed out by Spearman(1) that, since no psychological test is perfectly self-consistent, there is a constant tendency for test intercorrelations to be lower than they would be if they had been obtained from self-consistent test results. This lowering of test intercorrelations by errors of measurement in the tests was called by Spearman 'attenuation', and he showed how an estimate of the 'true' or 'unattenuated' coefficient of correlation might be obtained by the use of measures of self-consistency or 'reliability' obtained by correlating two independent measurements made on the same group of persons by each of the two tests in question.

The independent measurements necessary for this purpose may be obtained in any one of three ways, the choice between them being determined by considerations of applicability of the methods or their convenience for the particular tests used. The methods are:

(1) Repetition of the same tests either on the same day or on some different day.

(2) Division of the results of each test into two equivalent halves, e.g. by adding together separately the scores for the odd and for the even items in each test.

(3) Use of two pairs of tests, of which the two members of each pair are of the same form, the same kind of material and the same difficulty, but contain different material, e.g. the L and M forms of the new Terman Revised Intelligence Scale.

The method of correction is the same in all cases. Formulae of two

types were suggested by Spearman. In the present discussion, I shall confine myself to the first of these (which has been most generally used) and to the simplest case in which there are only two repetitions of each test (which also has been the usual condition in which the correction has been used).

The formula in question, as given in Spearman's first article, is

$$r_{AB}=\frac{\frac{1}{4}(r_{A_1B_1}+r_{A_2B_2}+r_{A_1B_2}+r_{A_2B_1})}{\sqrt{(r_{A_1A_2}\cdot r_{B_1B_2})}}.$$

Although the numerator of this equation is here given as the arithmetic mean of the four intercorrelations of the A's with the B's, the proof given by Spearman himself in a later article (2) and that supplied by Udny Yule (published as an appendix in a still later article by Spearman (3)) both lead to the geometric mean of these quantities. Spearman says in his first article that the geometric mean should be used if one test is much more accurate than the other. Although it will not make much difference which form is used if the r_{AB}'s are all about equal, the geometric mean seems to be the more correct and is always given by Spearman in his more recent writings. The formula then becomes

$$r_{AB}=\frac{\sqrt[4]{(r_{A_1B_1}\cdot r_{A_1B_2}\cdot r_{A_2B_1}\cdot r_{A_2B_2})}}{\sqrt{(r_{A_1A_2}\cdot r_{B_1B_2})}}.$$

In using this formula, it is necessary, of course, to know its standard error. This, as given by Kelley (4), is

$$\frac{r_{AB}}{2\sqrt{N}}\left(4{r_{AB}}^2+\frac{1}{{r_{AB}}^2}+\frac{1+r_{A_1A_2}+r_{B_1B_2}}{\sqrt{(r_{A_1B_1}\cdot r_{A_2B_2}\cdot r_{A_1B_2}\cdot r_{A_2B_1})}}+\frac{1}{{r_{A_1A_2}}^2}+\frac{1}{{r_{B_1B_2}}^2}-\frac{4}{r_{A_1A_2}}-\frac{4}{r_{B_1B_2}}-2\right)^{\frac{1}{2}}.$$

The simplest proof of the formula for correction for attenuation is the one first given by Udny Yule (3), which depends on the formula for the correction of sums. If a test X is administered on two occasions and the two sets of results are not perfectly correlated, we can consider the deviations of the first test as made up of two parts, one perfectly correlated with a corresponding part of the second test deviations and the other independent of the deviations in the second test. That is, $X_1=x+e_1$ and $X_2=x+e_2$, where x is the part common to the deviations of the first and second applications of the test and e_1 and e_2 are the parts of the deviations which show independence of each other and of x. Similarly, we may use the symbols y, f_1 and f_2 for the corresponding parts of the deviations of Y_1 and Y_2 respectively.

Now $\Sigma X_1 . X_2 = \Sigma (x+e_1)(x+e_2) = \Sigma x^2 + \Sigma x . e_1 + \Sigma x . e_2 + \Sigma e_1 . e_2$. But $\Sigma X_1 . X_2 = r_{X_1X_2} . \sigma_{X_1} . \sigma_{X_2}$, and similarly for the terms on the right-hand side of the equation. So the above equation may be written

$$r_{X_1X_2} . \sigma_{X_1} . \sigma_{X_2} = r_{xx} . \sigma_x^2 + r_{xe_1} . \sigma_x . \sigma_{e_1} + r_{xe_2} . \sigma_x . \sigma_{e_2} + r_{e_1e_2} . \sigma_{e_1} . \sigma_{e_2}.$$

But since $r_{xx}=1$, and by definition r_{xe_1}, r_{xe_2} and $r_{e_1e_2}$ are zero, it follows that

(*a*) $r_{X_1X_2} . \sigma_{X_1} . \sigma_{X_2} = \sigma_x^2$ and $r_{Y_1Y_2} . \sigma_{Y_1} . \sigma_{Y_2} = \sigma_y^2$.

By a similar argument:

(*b*) $r_{X_1Y_1} . \sigma_{X_1} . \sigma_{Y_1} = r_{X_2Y_2} . \sigma_{X_2} . \sigma_{Y_2} = r_{X_1Y_2} . \sigma_{X_1} . \sigma_{Y_2} = r_{X_2Y_1} . \sigma_{X_2} . \sigma_{Y_1}$
$= r_{xy} . \sigma_x . \sigma_y$.

From which we can easily derive the relationship: r_{xy} (the correlation between X and Y if there are no errors of measurement) is equal to

$$\sqrt{\frac{r_{X_1Y_1} . r_{X_2Y_2}}{r_{X_1X_2} . r_{Y_1Y_2}}} \quad \text{and also to} \quad \sqrt{\frac{r_{X_1Y_2} . r_{X_2Y_1}}{r_{X_1X_2} . r_{Y_1Y_2}}}.$$

Spearman's formula as given above is plainly got by taking the geometric mean of these two equations. This combination is sometimes legitimate and sometimes it is not; a better estimate of the unattenuated coefficient will, in certain cases, be obtained by taking it as equal to

$$\sqrt{\frac{r_{X_1Y_2} . r_{X_2Y_1}}{r_{X_1X_2} . r_{Y_1Y_2}}}.$$

It is important to notice, in the proof of this formula, that only the two equations (*a*) are mathematical identities. We have made these so by defining e_1 and e_2 as the parts of the deviations of the two testings of X which are independent of each other and of x. The four equations (*b*) are not mathematical identities, since we cannot define e_1 as the part of the deviations of X_1 which are uncorrelated with X_2 and also as the part uncorrelated with Y_1 and Y_2. This limitation is lost sight of if we begin by defining e_1, e_2, etc. as the 'errors of measurement' of the tests to which they belong.

Two consequences follow from this fact. First, the obvious consequence that the quantities, r_{ye_2}, etc., which are assumed to be zero in deriving the (*b*) set of equations will not be exactly zero but will at best approximate to zero within limits determined by the amount of their sampling error. This effect is, of course, allowed for in the calculation of the standard error of the final corrected coefficient. Its most striking consequence is that it may give the corrected coefficient a value greater than unity.

A second consequence is that some of the relationships indicated in the (*b*) set of equations may be altogether false if there is some special condition in the way of giving the tests which causes significant correlation between e_1 or e_2 and f_1 and f_2. There is an obvious way in which this might occur. Let us suppose that X_1 and Y_1 are administered on one day and that X_2 and Y_2 are administered on some other day, and that the mental function measured by X and Y is not constant in amount but varies for each subject from day to day. Then $r_{e_1f_1}$ and $r_{e_2f_2}$ will not be zero but will have a positive value which will depend on the extent to which any individual's amount of the function measured depends on the time of the measurement. $r_{X_1Y_1}$ and $r_{X_2Y_2}$ will, therefore, be greater than $r_{X_1Y_2}$ and $r_{X_2Y_1}$, and a true correction for attenuation cannot be obtained from $r_{X_1Y_1}$ and $r_{X_2Y_2}$, since the factor of contemporaneity which enhances these coefficients is not present in $r_{X_1X_2}$ and $r_{Y_1Y_2}$ in the denominator of the corrected coefficient (5). The formula to be used in correction for attenuation must, therefore, be $\sqrt{\frac{r_{X_1Y_2} \cdot r_{X_2Y_1}}{r_{X_1X_2} \cdot r_{Y_1Y_2}}}$.

Since this use of the two 'different time' test intercorrelations sufficiently determines the value of the corrected coefficient (whether or not $r_{X_1Y_1}$ and $r_{X_2Y_2}$ are greater than $r_{X_1Y_2}$ and $r_{X_2Y_1}$), it is suggested that this formula should always be used when X_1 and Y_1 are measured together at a time different from that of X_2 and Y_2. Otherwise, the use of the formula involving all four intercorrelations seems unobjectionable. It should be noticed that the above formula gives the correlation which would have been found between X and Y if neither errors of measurement nor fluctuation of the function measured had been present, i.e. the effect of function fluctuation in lowering the coefficient of correlation between X and Y has been eliminated as well as the effect of errors of measurement. The use of the ordinary formula would, on the other hand, have given a spuriously high value of the corrected coefficient if there had been day-to-day fluctuation of the mental function measured, and it might have given a corrected coefficient significantly greater than unity. It seems not unlikely that when correlation coefficients corrected for attenuation have been found to be significantly greater than unity, the error has arisen in this way. Fluctuation of functions measured by tests is apparently not uncommon when these functions are character traits (5). It appears to be absent in intelligence, and this may apply also to any measurement of an ability.

It seems that this same cause might result in a spurious lowering of the corrected coefficient under another condition of experiment. Let us

suppose that test X was applied on one day and test Y on another, and the method of equivalent halves was used for correcting attenuation. Then $r_{X_1X_2}$ and $r_{Y_1Y_2}$ would be increased (if function fluctuation were present) by having been measured at the same time, but this increase would not be found in any of the four correlations between X and Y. Thus the effect of function fluctuation will be to increase the denominator of the correction formula while leaving unaffected the numerator. There seems to be no easy way of evading this difficulty, so that this arrangement of the tests would appear to be one to be avoided unless it is quite certain that we are dealing with a function which does not show day-to-day fluctuation.

II. Objections to correction for attenuation and difficulties in its use

The attitude of psychological testers towards correction for attenuation has been very variable. A few seem to have used it generally as a routine treatment of correlation coefficients; a few have rejected the method either because they regarded it as inappropriate to psychological data or because they were rendered uneasy by an abnormality of corrected coefficients which was criticized by Karl Pearson; probably the largest number have admitted the necessity for correction in theory but ignored it in practice. The suggestion I am putting forward in the present paper is that correction for attenuation is essential in a certain number of problems for which correlation coefficients are used, that it is unnecessary in a much larger number of problems, and definitely incorrect for certain purposes.

Before trying to distinguish the purposes for which correction for attenuation is necessary, is unnecessary, and is wholly inappropriate, it may be as well to mention the objections which have been urged against the use of the correction for any purpose. Karl Pearson(6) considered that the method was sufficiently condemned by the fact that it sometimes led to the absurd result of giving a value of r which was greater than one. This, as Spearman has pointed out, is simply due to the fact that the sampling error of a coefficient corrected for attenuation is greater than would be that of an uncorrected coefficient of the same size. It results from the fact that the argument for the formula assumes that the intercorrelations of the 'errors of measurement' are zero, whereas they may, in fact, have values positive or negative differing from zero within the limits imposed by their own probable errors. It would certainly be more pleasant to have a method for correcting for attenuation

which had unity as its theoretical maximum for the corrected coefficient, and it is possible that such a method may be devised some time. Until it is, the possibility of getting a corrected r which is greater than unity is not a very serious defect of the method if it is remembered that the estimate of r given by the method is (like every other statistic) limited in its accuracy by its sampling error. If the corrected coefficient is correctly calculated from appropriate data, it will not be found to be significantly greater than unity.

While this is a sound (and, I think, adequate) defence against the charge that the occurrence of values of the correlation coefficient greater than unity is a fatal objection to the use of correction for attenuation, I have, nevertheless, the impression (which I cannot support by exact figures) that coefficients corrected for attenuation come out as greater than unity rather more often than they should. Since investigators frequently give corrected coefficients without giving the raw coefficients on which they are based or saying what formula was used for correction, it is not possible to be certain in what way this tendency to over-correction (if it exists) has arisen. If the formula with the arithmetic mean were used instead of that with the geometric mean, this would cause a slight over-correction, since the geometric mean of a number of unequal figures is always less than their arithmetic mean. The effect would be small unless the correlations of the numerator were very unequal. A more likely source of over-correction is the use of $r_{A_1B_1}$ and $r_{A_2B_2}$ under the conditions already mentioned in which function fluctuation makes their use invalid.

The use of correction for attenuation has also been objected to not on mathematical but on psychological grounds. It is sometimes said that the inconsistencies between two successive administrations of the same test to a group are not 'errors' but are all psychologically caused; a subject who was attentive when answering an item in one administration and got it right may make a mistake in the same item when the test is repeated because he is distracted, and so on. Variations in attention are real psychological events and not accidents. This objection has, however, no relevance to the question of the legitimacy of correction for attenuation, since this method does not assume that the so-called 'errors' are accidental. All that is necessary for the justification of the method is that the same test twice applied to the same group gives two sets of results that are not exactly the same. The individual deviations can then be regarded as the sum of two parts, one part showing perfect correlation between the two applications of the test and the other

showing no correlation. The uncorrelated elements in the deviations need not be accidental; they are treated as 'error' because they are irrelevant to the purpose in hand, which is the assessment of the relative abilities of the tested individuals in whatever capacity the test is designed to measure. For other purposes, these fluctuations may be relevant; their irrelevance (the character in virtue of which they are treated as 'error') is not absolute; it is only relative to the purpose of the test.

In addition to objections to the attenuation correction in principle, there are certain particular difficulties in its application which depend on the fact that particular test results may not conform to the conditions that are assumed in the theory of the method. One such difficulty has already been discussed, that caused by day-to-day fluctuation of the function measured.

Another difficulty arises from the fact that if $r_{X_1X_2}$ and $r_{Y_1Y_2}$ are obtained in each case by repetition of the same test, it is possible that the correlation between them may be increased by the subject remembering his previous answers and repeating previous successes and failures in this way. This factor will lead to underestimation of the corrected coefficient. Since there is no reason for using correction for attenuation unless an estimate of the absolute value of the correlation coefficient is required, under-correction can no more be tolerated than over-correction. This memory effect will become less as time between the two testings is increased and may become negligible if sufficient time is allowed to elapse, partly because the subject's memory of previous responses rapidly fades, and partly because (in the case of children) increase of chronological age will make the critical questions fall in a new part of the test on the retesting. To some kinds of test, the objection does not apply at all. It is, for example, very improbable that memory would cause any intercorrelation between errors in Seashore's pitch-discrimination test even if the repetition were given immediately after the first application.

This consideration leads us to prefer for many tests the alternative of dividing it into two equivalent halves or of giving another test in the same form but with different material. If either of these methods is used, it is obviously necessary that equivalence should be as close as possible, while similarity must not be so close that the first objection applies. If the two similar tests do not resemble each other sufficiently closely, there may be expected to be a considerable specific factor in each which will lower the reliability coefficients and so enhance the value of the coefficient of correlation corrected for attenuation. Since in every case where a reliability coefficient has been calculated from similar but not

identical tests it is to be expected that there will be at least a small specific factor, we may notice that this is not a serious objection to the use of such material for correction for attenuation. Let us suppose that the deviations of each test result are made up of a factor common to the other test of the same kind, of a specific factor, and of an error of measurement. Thus the deviations of X_1, X_2, Y_1 and Y_2 can be represented as $(x+s_1+e_1)$, $(x+s_2+e_2)$, $(y+s_3+f_1)$ and $(y+s_4+f_2)$, respectively. Now if the corrected coefficient had been obtained by merely repeating X_1 and Y_1, its value would have been that of the correlation between $(x+s_1)$ and $(y+s_3)$. If, however, the corrected coefficient is calculated from the two pairs of non-identical tests X_1 and X_2 and Y_1 and Y_2, the value obtained is that of the correlation between x and y. This will be larger than the value obtained by the other method, but it will be in no sense a spurious enhancement of the correction. This value also has a definite and legitimate meaning.

There is, however, a more serious objection to the use of the self-correlation of two halves of the test as a measure of the test reliability and as a means of correcting for attenuation, which applies to most (though not to all) modern group tests of intelligence. These are very generally constructed so that either the test as a whole or each section of it has a time limit such that most testees do not have time to complete all the test items. This method of construction has obvious practical advantages, but it has the unfortunate effect that the two equivalent halves into which the test items can be divided do not correspond to two independent applications of equivalent tests.

Let us suppose, for example, that a particular section of a test contains twenty items and that a particular subject manages to do twelve of them in the time limit allowed. On a retesting, he might do rather more or rather less than this, and such random variations of speed are part of what makes the self-correlation of the test fall below unity and attenuates its correlations with all other tests. It does not, however, reduce the correlation between the two halves of the test formed in such a way as that of taking the odd and the even test items, for six of each of these have been done in the time allowed for the section. If the number of test items completed were the only factor in determining individual differences in score, the 'reliability' as measured by the intercorrelation of equivalent halves would be approximately unity. Actually, scores are also determined by the number of mistakes and these will be distributed over the two equivalent halves much as they would be between two independent tests, and will cause the half-half correlation to fall short of

unity, but to an extent which may considerably over-estimate the real reliability of the test.

The difficulty here is that the self-correlation of independent testings is determined by the variability of two factors—speed and correctness. Of these, correctness alone determines the intercorrelation of equivalent halves. Therefore the reliability is over-estimated to an extent which depends on the relative importance of speed and of correctness as determinants of individual differences in score. Any correction for attenuation obtained from such a 'reliability' coefficient will suffer from serious under-correction. The equivalent-halves method cannot validly be used for determining the reliability of a test which has a time limit either for the whole test or for its sections. The originators of tests of this kind have sometimes published 'reliability' coefficients obtained in this way. It is apparent both in theory and in practice that these estimates of reliability are dangerously over-optimistic.

The fact that each subject has practised the first test by the time he does the second may also cause a difference between the two sets of responses which, if the effects of the practice are different in different subjects, will lower the correlation between X_1 and X_2 and between Y_1 and Y_2 and so lead to over-correction for attenuation. Kelley (4) has suggested that this difficulty may be got over by giving sufficient fore-exercise for each test to establish the appropriate attitude and to give familiarity with its structural features.

Thus it may be seen that there are certain conditions to be observed in the collection of data for correction for attenuation. With the exercise of sufficient care, data for most tests can be made appropriate for this operation. Sometimes, probably, the difficulties are insuperable. I know, for example, of no performance scale of intelligence that can be repeated without memory largely influencing the retest result, or that can be split into two equivalent parts, or that can be considered effectively equivalent to any other performance scale. Valid correction for attenuation would be very difficult with such material. More generally the difficulties may be overcome by care in planning the design of the experiment.

III. Conditions in which coefficients of correlation should and should not be corrected for attenuation

If we have agreed that correction for attenuation is a legitimate operation when applied to correlation coefficients that have been obtained under appropriate conditions, it does not follow that this is an operation

which it is always either necessary or desirable to apply in psychological research. Whether correlation coefficients should or should not be corrected for attenuation depends on the purpose for which they are required.

Let us consider then the purposes for which we may be calculating a coefficient of correlation:

(1) We may want to know how far success in one test is predictive of success in another test. It is, for example, a common problem in educational psychology to discover how successfully the results of school entrance examinations predict subsequent school success. Here the answer to our problem is given by the correlation coefficient uncorrected for attenuation and the corrected coefficient would be definitely wrong. We want to know what is the predictive value of the actual entrance examination, not what would be the predictive value of an ideal entrance examination with perfect self-consistency. A prophet must be judged by his success in foretelling the future under the actual conditions in which he has to prophesy, not by an estimate of how well he would foretell the future if these conditions were improved. Lack of self-consistency is one of the conditions under which the entrance examination makes its prediction of future success, and to eliminate the effect of this condition would be to give a falsely favourable estimate of its predictive success.

This may be stated more generally. Whenever we want to determine the amount of correlation between test measurements (and not the amount of correlation between the abilities or other mental functions measured by the tests), then the attenuated coefficient must be used.

(2) It very commonly happens that we work out a correlation coefficient in order to determine whether or not two measurable mental characteristics are causally connected. Let us suppose that we want to test the hypothesis that skills depending on hand-eye co-ordination are easier to acquire if the eyes are wider apart. One way of solving this problem would be to measure the distance between the eyes of a large group of persons and to test their skill in a number of such performances as throwing darts, catching balls, etc., and to work out the correlation coefficient between eye distance and each of these skills. Causal connexion of some kind or other is indicated by the discovery of a significant correlation, and if our purpose is only to discover which (if any) of these skills are causally dependent on eye distance, we need only to determine which of the correlations are significant. The fact of causal connexion will be made no clearer by knowing how great is the correlation between eye distance and the capacity in question. If we are interested only in the fact of causal connexion, there is no possible advantage in working out

the coefficient as corrected for attenuation; if the corrected coefficient is significant, then so also will be the uncorrected coefficient. The attenuated coefficient is, therefore, to be preferred, since it requires much less labour to determine and gives the required information equally well.

Generally, if our object is to demonstrate the existence of correlation between A and B and not to determine its amount, then the attenuated coefficient serves our purpose as well as the coefficient corrected for attenuation. The uncorrected coefficient is, therefore, the better one to use, since the additional labour of correcting for attenuation serves no useful purpose.

(3) It also frequently happens that we require a correlation coefficient in order to perform some operation with it whose validity does not depend on correction for attenuation. We may, for example, wish to discover whether a group factor of 'mathematical ability' underlies the ability in different mathematical subjects. We test a group of school children in geometry and algebra and try to determine by means of Spearman's 'tetrad equation' whether the correlation between these two abilities is greater than would be due to the dependence of both on general intelligence or G. We obtain the group scores in two tests which have no other common factor than G (let us say a verbal and a perceptual test of intelligence). Then if the tetrad difference $r_{GA}.r_{VP}-r_{GV}.r_{AP}$ is significantly greater than zero, we can conclude that ability in geometry and algebra depend on a common factor which is not G. Now Spearman has shown that a tetrad relationship which holds between coefficients corrected for attenuation will hold also between coefficients uncorrected for attenuation. Here again, therefore, it is unnecessary to correct for attenuation, and in problems of this type it is better to save labour by not doing so.

To put this in general terms. If we wish to use a correlation coefficient to demonstrate a relationship which holds equally between corrected and uncorrected coefficients, correction for attenuation is unnecessary and should, therefore, not be used.

(4) The cases so far discussed cover the greater number of the actual problems for which coefficients of correlation are required in psychological research work. There remains the case in which we wish to know the degree of correlation, not between the test results but between the abilities measured by the tests. Sometimes the problem is of the correlation between character traits or between characteristics of perception which are not of the nature of abilities. I shall, therefore, use the general term 'mental function' for what is measured by tests, meaning to include

in this term such things as abilities, character traits, perceptual characteristics (such as phenomenal regression) and so on. Where the problem is of the absolute amount of correlation between mental functions, the correction for attenuation must be used. If any psychologist feels so strongly the objections to correction for attenuation which have already been discussed that he is unable to use this correction, then he cannot use correlation coefficients for the solution of the present problem. The use of the coefficient uncorrected for attenuation as a measure of the amount of correlation between the mental functions measured by two tests is wrong and wholly unjustifiable. All that the uncorrected coefficient tells us is that the required coefficient of correlation lies somewhere between the value indicated and unity.

We might, for example, want to know to what extent ability to pass school entrance examinations depended on the same individual characteristics of children as did later school success. This would be a wholly different problem from that discussed in (1), although it would start from the same primary data. For answering this question, the coefficient of correlation must be corrected for attenuation. The uncorrected coefficient would under-estimate to an unknown extent the amount of what was common to the ability to pass the entrance examination and the ability to succeed later at school.

We certainly sometimes want to obtain information of this kind. Probably we want it less often than the psychological investigator is inclined to imagine, since there is a tendency to suppose that the coefficient so obtained gives more information than it actually does. The correlation coefficient is often spoken of as if it gave a measure of the strength of the causal connexion between two variates. In reality the information it gives is of a less fundamental character than this. To quote Fisher(8): "It [the correlation coefficient] tells us the relative importance of the factors which act alike upon the [variates in question] compared to the totality of factors at work." In other words, the correlation coefficient can be represented as the fraction $X/(X+Y)$, in which X represents the variance of the factor which is common to both tests and Y represents that part of the total variance which is specific to each test.

The magnitude of the correlation will thus depend on the relative importance of the factors specific to each test in the particular population from which the sample investigated has been drawn (i.e. on the size of Y). Inequalities of teaching and of school attendance will, for example, probably be some of the factors causing individual differences in later school success which are uncorrelated with ability to pass the school

entrance examination. If we measure the correlation in a sample drawn from a population with great inequalities either of teaching or of attendance, then the correlation coefficient will be correspondingly smaller. If these inequalities are less the correlation will be greater. If these are the only factors other than initial scholastic ability which influence school success, then if there are no inequalities in them the correlation will be unity. In all cases the correlation coefficient is a measure not of the absolute importance of initial scholastic ability in determining later school success but of its importance relative to various other factors. If, as certainly sometimes happens, it is important for us to have this information, then the correlation coefficient corrected for attenuation is what we require.

We may thus see that in the practical psychological investigations for which we require correlation coefficients, it may happen (although not, I think, very commonly) that a coefficient corrected for attenuation is the only one that will give us a correct answer to our question. In the majority of our inquiries, a corrected or uncorrected coefficient will do equally well. Sometimes (although again not very commonly) an uncorrected coefficient is the only one that answers our question correctly. These facts (and especially the last) are obscured by a common habit of referring to the coefficient corrected for attenuation as the 'true correlation', with the obvious implication that otherwise the coefficients are untrue. The previous argument shows that the coefficient uncorrected for attenuation may be as true as or truer than the corrected coefficient. It is always a perfectly true measure of the degree of correlation between test measurements; it only becomes untrue when it is taken also to be a measure of the degree of correlation between the functions measured by the tests. I suggest that this can be best expressed by abandoning the term 'true' correlation for the coefficient corrected for attenuation, and substituting the term 'function' correlation. The coefficient uncorrected for attenuation can similarly be called the 'measurements' coefficient of correlation. It is not merely an incorrect estimate of the correlation between the functions measured by the two tests; it is a correct estimate (i.e. correct within the limits of its own sampling error) of the correlation between the measurements given by the two tests. Both correlations are, for different purposes, equally 'true'.

There is one other possibility which, I think, deserves consideration. Let us suppose that we wish to find out how well a newly devised intelligence test (A) predicts the general ability measured by some established test of intelligence (B), let us say, the *Stanford Revision*. Here we have a

problem intermediate between that for which the 'function' correlation and the 'measurements' correlation is required. The coefficient of correlation corrected for attenuation would over-estimate the predictive value of the test, since it would tell us not how well the test does in fact diagnose intelligence as measured by the standard test but how well it would diagnose it if our new test were free from unreliability (i.e. under conditions better than those in which the diagnosis must in fact be made). If entirely uncorrected for attenuation, however, the coefficient would under-estimate the diagnostic value of the test, since it would be including, as an imperfection of the test, the unreliability of the standard. Here what we want to know is how well the measurement of our new test correlates with the function measured by the standard test; the effect of errors of measurement in the standard must be eliminated but not the effect of those in the test which is under investigation. By the same kind of argument as is used for the full correction for attenuation, it can be shown that the correlation here required is given by the formula

$$\sqrt{\frac{r_{AB_1} \cdot r_{AB_2}}{r_{B_1B_2}}}.$$

I suggest that this should be called the 'measurement-function' coefficient of correlation. Its general use is to determine how well the measurements in one test diagnose the function measured (validly but not with perfect reliability) by some other test.

That the use of a correlation coefficient (or any other statistic) should be directed towards the obtaining of an answer to a particular question may seem a point too obvious to be worth emphasizing. It is, however, surprisingly often ignored. The research worker in educational psychology too often accumulates a series of arrays of test results, and works out correlations between them, sometimes correcting for attenuation and partialling out any variates for which he has the necessary test data, all with no clearer object than the pious hope that effort in a procedure at once so respectable and so laborious must lead to some kind of useful result. It is not surprising that this hope is very commonly disappointed.

IV. Partial correlation

Another correction commonly applied to correlation coefficients is for eliminating the effect of the partial correlation of either or both variates with some other variate. We have, let us say, given tests in two school subjects to a group of children from different forms in a school. Those

from higher forms will, on the whole, do better in both tests because they have studied the subjects longer and at a higher standard than the others. This fact will make the correlation higher than it would have been if all the children had been of the same school standing. We may wish to know what the correlation between the tests would have been if all the children had been of the same school standing. This is determined by the use of Udny Yule's partial correlation formula:

$$r_{AB(C)} = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{(1 - r_{AC}^2)\,(1 - r_{BC}^2)}},$$

in which r_{AC} and r_{BC} are the correlations of the two tests with school standing and $r_{AB(C)}$ is the required correlation between the two tests if school standing is held constant (that is, in a population homogeneous with respect to school standing). Generally the corrected coefficient will be less than it was before if the eliminated variate affects both A and B in the same direction, but the correction will increase the coefficient if the eliminated variate affects one of the correlated variates but not the other, or if it affects both in opposite directions. In such an inquiry as this it may, of course, happen that we want not only to eliminate the effect of differences in school standing but also in chronological age, or in intelligence or in both.

Let us suppose that we wish to eliminate the effect on r_{AB} of differences in intelligence. We are then faced by a very serious limitation to the use of this formula which is not always realized by psychologists who want to work out a partial correlation; the formula is only valid if the eliminated variable C can be measured without error. No test result fulfils this condition and, therefore, this formula can never be used to partial out any test result unless all the correlations used have been corrected for attenuation. It seems that the only two things we can measure without error in educational psychology are chronological age and school standing. Neither of these will, it is true, be normally distributed in any results we are likely to get in a practical investigation, but M. S. Bartlett (7) has shown that normal distribution is not necessary in the eliminated variate in a partial correlation. The partial correlation coefficient can, therefore, be used as it stands for eliminating the effect of chronological age and of school standing if these are irrelevant factors for the purpose for which we require the coefficient. Otherwise its use is illegitimate and will lead to misleading results.

When we say that the use of the ordinary partial correlation formula is misleading when the measurements of the eliminated variate are

subject to error, what is meant is not merely that its use gives an incorrect estimate of $r_{AB(C)}$. That is obvious, and, for many purposes, may be unimportant. A much more serious objection to the use of this formula in this case is that it cannot be relied upon to provide accurate information as to whether A and B are correlated at all in a population homogeneous with respect to C. That, at least, we want to know when we work out a partial correlation coefficient. As will be shown later, however, the ordinary formula may (and normally will) give us a value of $r_{AB(C)}$ differing from zero if C is subject to errors of measurement when, if C had been measured without these errors, $r_{AB(C)}$ would have been zero. The ordinary formula cannot, therefore, in this case, even be relied on to tell us whether A and B are correlated when C is held constant. In other words, if C is appreciably affected by errors of measurement, the application of the ordinary formula without any correction for these errors of measurement is of no use whatever.

An obvious way out of this difficulty is to obtain duplicate testings of A, of B and of C, and to correct all the coefficients for attenuation by use of Spearman's formula. This is legitimate but laborious. Occasionally it is not possible, since some tests (e.g. some of the performance scales of intelligence) cannot be repeated and cannot be split up into equivalent halves. If this procedure of complete correction for attenuation is carried out, it leads to the discovery of the 'function' correlation between A and B when C is constant. This is generally satisfactory, but it may sometimes happen that it is the 'measurements' correlation between A and B that is required.

I suggest that when the 'measurements' correlation is needed or when it will serve our purpose just as well as the 'function' coefficient, a method of dealing with the problem of partial correlation less radical than that of complete correction for attenuation should be adopted. The self-correlation of C must be known, so there must be double measurement of C but not of A or B. The required partial correlation between A and B is then given by the formula

$$\frac{r_{AB}.r_{C_1C_2}-\sqrt{(r_{AC_1}.r_{AC_2}.r_{BC_1}.r_{BC_2})}}{\sqrt{\{(r_{C_1C_2}-r_{AC_1}.r_{AC_2})\,(r_{C_1C_2}-r_{BC_1}.r_{BC_2})\}}}.$$

The proof of this formula is given in the Appendix.

The relationship of this formula to the usual formula for partial correlation can be made most clear if we remember that r_{AC_1} will be approximately the same as r_{AC_2} and that r_{BC_1} will be approximately the same as

r_{BC_2}. So $\sqrt{(r_{AC_1}r_{AC_2})}$ and $\sqrt{(r_{BC_1}r_{BC_2})}$ may be written as r_{AC} and r_{BC} respectively. The formula then becomes

$$\frac{r_{AB}.r_{C_1C_2}-r_{AC}.r_{BC}}{\sqrt{\{(r_{C_1C_2}-r_{AC}^2)\,(r_{C_1C_2}-r_{BC}^2)\}}},$$

from which it is obvious that if $r_{C_1C_2}$ were unity (i.e. if there were no error of measurement in test C), the formula would reduce to the usual form for partial correlation.

The purpose of the formula we are at present discussing is to give the partial correlation between the *measurements* of A and B in a population homogeneous with respect to C. It is suggested that for purpose (1) (of the earlier discussion, on p. 392), it is the only right partial correlation formula to use if C has errors of measurement, that for purposes (2) and (3) it is preferable to the method of complete correction for attenuation since it is less laborious, while for purpose (4) it is inadmissible and the method of complete correction for attenuation must be used.

It might be supposed that since it has been argued that the total correlation coefficient uncorrected for attenuation has a definite meaning (that of correlation between test results) and is the proper measure of correlation for some purposes, the partial correlation coefficient calculated from the usual formula when the eliminated variable is subject to uncorrected errors of measurement might be justified for certain purposes by a similar argument. Has not this coefficient also a definite statistical meaning? Certainly it has, but it is not a meaning that is of any practical use to us. It is a measure of the correlation that would exist between A and B in a population in which the test scores in test C were constant. I do not think that there is any conceivable inquiry in which this would be a useful piece of information. Perhaps the greatest use of the attenuated coefficient of correlation is the second use mentioned above—to tell us whether the two tests in question are correlated. If the attenuated correlation coefficient is significant, then the functions measured by the two tests are correlated. Correction for attenuation will give us a better estimate of how much they are correlated (if we require to know this) but it will give us no more certain indication of the fact of correlation.

There is no similar justification for the use of the uncorrected partial correlation coefficient. Let us suppose that two abilities A and B are correlated in a population heterogeneous with respect to C, and that our reason for using the partial correlation method is to discover whether or not they are correlated in a population homogeneous with respect to C. We cannot infer this from observing whether or not they are correlated

in a population homogeneous with respect to C scores, since they may be correlated in a population homogeneous with respect to C scores although uncorrelated in a population homogeneous with respect to the function C itself. This is clear if we consider the case that A and B are uncorrelated in a population with C constant but are correlated in a population heterogeneous with respect to C. Then, by the formula last given, we know that $r_{AB}.r_{C_1C_2}-r_{AC}.r_{BC}=0$. Since, however, $r_{C_1C_2}$ is less than one (if the test of C is not perfectly self-correlated), it follows that $r_{AB}-r_{AC}.r_{BC}$ must be greater than zero. It follows that the usual formula will make A and B appear to be correlated when C is held constant although, in fact, they are uncorrelated. Obviously it is also theoretically possible (although less likely to happen in practice) that this formula will make A and B appear to be uncorrelated when they really are correlated.

The usual formula is, therefore, to be condemned as a method of obtaining partial correlations whenever the variate partialled out is subject to appreciable errors of measurement. This means that it is to be condemned for most purposes for which it is likely to be required in psychology. It is necessary, either that A, B and C should be measured twice and all coefficients corrected for attenuation, or that C alone should be measured twice and that the formula I have suggested should be used. The first procedure will give the 'function correlation' between A and B, and the second will give the 'measurements correlation' between A and B, both with C held constant. Which procedure will be adopted must depend on the purpose for which the coefficient is required, as has been discussed earlier.

V. Summary

Correction for attenuation is a justifiable operation when properly applied to appropriate data, although it is not always (or most commonly) necessary.

Certain precautions are necessary when applying it to coefficients of correlation between mental functions which show day-to-day fluctuation.

The required estimate of reliability may be obtained: (A) by repeating the same test, (B) by splitting it into two equivalent halves, or, (C) by repeating a test of equivalent difficulty. (A) is not appropriate if similarity of responses in test and retest might be caused by memory. (B) is not appropriate if the test has a time limit.

Correction for attenuation is wrong if what is required is the correlation between two test scores. It is not necessary if we require only to know that two tested functions are correlated, or if we want to use the

coefficients for the investigation of some relationship which holds equally between attenuated and unattenuated coefficients. It is necessary only when we want to know the absolute value of the correlation coefficient between the two tested functions. Consideration of the meaning of the correlation coefficient suggests that this knowledge is probably less often of importance than is commonly supposed.

The usual formula for correction for partial correlation is not valid if the eliminated variate is subject to errors of measurement. This severely limits its usefulness in psychological investigation. A formula is suggested for the valid application of the correction for partial correlation when it is not desired (or not possible) to correct all the coefficients used for attenuation.

APPENDIX

Proof of the formula for obtaining the 'measurements' correlation between A and B, with C held constant, C being itself subject to errors of measurement.

Four test measurements must be made: A, B, C_1 and C_2 (i.e. test and retest of C).

Let a and b be the elements in the deviations specific to A and B respectively (including their errors of measurement).

Let n be the element in the deviations of A and B which is common to them but not to C.

Let m be the element in the deviations of C which is found also in A and B with weights represented by pm and qm, respectively.

Let c_1 and c_2 be the elements in the deviations of C_1 and C_2, respectively, which are not common to both measurements of C (i.e. their respective errors of measurement).

Then the deviations of A are compounded of $a+pm+n$.

" " " B " " $b+qm+n$.

" " " C_1 " " c_1+m.

" " " C_2 " " c_2+m.

The partial correlation required is the correlation between A and B with m held constant, that is, $r_{(a+n)\,(b+n)\cdot}$

Now $$r_{(a+n)(b+n)}\cdot\sigma_{(a+n)}\cdot\sigma_{(b+n)}=r_{ab}\sigma_a\sigma_b+r_{an}\sigma_a\sigma_n+r_{bn}\sigma_b\sigma_n+r_{nn}\sigma_n^2$$
$$=\sigma_n^2$$

(since, by hypothesis, a, b and n are uncorrelated, while $r_{nn}=1$).

So $$r_{AB(m)}=\sigma_n^2/[\sigma_{(a+n)}\cdot\sigma_{(b+n)}]. \quad \text{.....(1)}$$

Similarly, $$r_{AC_1}\sigma_A\cdot\sigma_{C_1}=p\cdot r_{mm}\sigma_m^2=p\cdot\sigma_m^2$$

and $$r_{AC_2}\sigma_A\cdot\sigma_{C_2}=p\cdot\sigma_m^2,$$

so $$r_{AC_1}r_{AC_2}\sigma_A^2\cdot\sigma_{C_1}\cdot\sigma_{C_2}=p^2\cdot\sigma_m^4; \quad \text{.....(2)}$$

similarly, $$r_{BC_1}r_{BC_2}\sigma_B^2\cdot\sigma_{C_1}\cdot\sigma_{C_2}=q^2\cdot\sigma_m^4. \quad \text{.....(3)}$$

Also $$r_{C_1C_2}\sigma_{C_1}\sigma_{C_2}=r_{c_1c_2}\sigma_{c_1}\sigma_{c_2}+r_{c_1m}\sigma_{c_1}\sigma_m+r_{c_2m}\sigma_{c_2}\sigma_m+r_{mm}\sigma_m{}^2$$
$$=\sigma_m{}^2 \qquad \text{.....(4)}$$

(since, by hypothesis, c_1, c_2 and m are uncorrelated).

In the same way, we find that

$$r_{AB}\sigma_A\sigma_B=\sigma_n{}^2+pq.\sigma_m{}^2, \qquad \text{.....(5)}$$

so $$\sigma_n{}^2=r_{AB}\sigma_A\sigma_B-pq.\sigma_m{}^2.$$

Substituting this in (1), we get

$$r_{AB(m)}=(r_{AB}\sigma_A\sigma_B-pq.\sigma_m{}^2)/[\sigma_{(a+n)}\sigma_{(b+n)}]. \qquad \text{.....(6)}$$

Multiplying (2) and (3) and taking the square root:

$$pq.\sigma_m{}^4=\sqrt{(r_{AC_1}.r_{AC_2}.r_{BC_1}.r_{BC_2})}.\sigma_A\sigma_B\sigma_{C_1}\sigma_{C_2}.$$

Dividing by (4), this becomes

$$pq.\sigma_m{}^2=\sqrt{(r_{AC_1}.r_{AC_2}.r_{BC_1}.r_{BC_2})}.\sigma_A\sigma_B/r_{C_1C_2}. \qquad \text{.....(7)}$$

By the ordinary formula for summing independent variances

$$\sigma_A{}^2=\sigma_{(a+n)}{}^2+p^2\sigma_m{}^2.$$

So $$\sigma_{(a+n)}{}^2=\sigma_A{}^2-p^2\sigma_m{}^2 \text{ and } \sigma_{(b+n)}{}^2=\sigma_B{}^2-q^2\sigma_m{}^2.$$

Therefore $$\sigma_{(a+n)}.\sigma_{(b+n)}=\sqrt{\{(\sigma_A{}^2-p^2\sigma_m{}^2)(\sigma_B{}^2-q^2\sigma_m{}^2)\}}. \qquad \text{.....(8)}$$

If we divide each side of (2) and of (3) by (4), we get

$$p^2\sigma_m{}^2=r_{AC_1}.r_{AC_2}.\sigma_A{}^2/r_{C_1C_2} \text{ and } q^2\sigma_m{}^2=r_{BC_1}.r_{BC_2}.\sigma_B{}^2/r_{C_1C_2}.$$

Substituting these values in (8):

$$\sigma_{(a+n)}.\sigma_{(b+n)}=\sigma_A\sigma_B\sqrt{\{(1-r_{AC_1}r_{AC_2}/r_{C_1C_2})(1-r_{BC_1}r_{BC_2}/r_{C_1C_2})\}}. \qquad \text{.....(9)}$$

Substituting (7) and (9) in (6), and multiplying numerator and denominator by $r_{C_1C_2}/\sigma_A.\sigma_B$:

$$r_{AB(m)}=\frac{r_{AB}.r_{C_1C_2}-\sqrt{(r_{AC_1}.r_{AC_2}.r_{BC_1}.r_{BC_2})}}{\sqrt{\{(r_{C_1C_2}-r_{AC_1}r_{AC_2})(r_{C_1C_2}-r_{BC_1}r_{BC_2})\}}}.$$

This is the formula for the partial correlation between the measurements of A and B when the variate partialled out is subject to errors of measurement.

If there were no errors of measurement in C, $r_{C_1C_2}$ would be unity, and

$$r_{AC_1}=r_{AC_2}=r_{AC} \text{ and } r_{BC_1}=r_{BC_2}=r_{BC},$$

so the above formula would then be reduced to the usual form:

$$r_{AB(C)}=\frac{r_{AB}-r_{AC}r_{BC}}{\sqrt{\{(1-r_{AC}{}^2)(1-r_{BC}{}^2)\}}}.$$

REFERENCES

(1) SPEARMAN, C. (1904). "'General Intelligence' objectively determined and measured." *Amer. J. Psychol.* XV, 201.
(2) —— (1907). "Demonstration of formulae for true measurement of correlation." *Amer. J. Psychol.* XVIII, 161.
(3) —— (1910). "Correlation calculated from faulty data." *Brit. J. Psychol.* III, 271.
(4) KELLEY, T. L. (1923). *Statistical Method.* New York.
(5) THOULESS, R. H. (1936). "Test unreliability and function fluctuation." *Brit. J. Psychol.* XXVI, 325.
(6) PEARSON, K. (1904). "On the laws of inheritance in man" (addendum on p. 160). *Biometrika,* III, 131.
(7) BARTLETT, M. S. (1932). "On the theory of statistical regression." *Proc. roy. Soc. Edinb.* LIII, 260.
(8) FISHER, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh.

(*Manuscript received* 24 *May* 1938)

THE GENERAL FACTOR IN CORRELATIONS BETWEEN PERSONS

By M. DAVIES

From the Psychological Laboratory, University College

I. *Problem* (pp. 404–405).
II. *Methods* (pp. 405–409).
III. *Data* (pp. 409–412).
IV. *Results* (pp. 412–416).
V. *Summary of conclusions* (pp. 417–420).
A. *As regards methods* (p. 417).
B. *As regards results* (pp. 417–420).
References (p. 420).

I. Problem

It is a familiar fact that, on correlating mental tests, the majority of the correlations prove to be positive, and that a single general factor will account for by far the larger proportion of the variance. It is therefore natural to enquire whether the same remarkable result will reappear when we correlate, not tests, but persons. To this question two strongly conflicting answers have been given.

Most earlier investigators have apparently assumed that the situation should be much the same in both cases. This view has been explicitly maintained by Burt (20) and Dewar (25); and is implied by the work of Bulley (see (9)), Beebe-Center (10), Pritchard (12), Shakespeare (19), and many others. Stephenson, on the other hand, in a recent series of papers (15,16,17) has vigorously contested this point of view. He insists upon a sharp opposition between the two methods of approach. Where correlations between tests are concerned, he still upholds the doctrine of a single general factor; but, where correlations between persons have been calculated, he finds, in nearly every case, not one general factor but two group factors, each limited to a particular group of persons, and together producing a pair of mutually exclusive 'types'.

If this conclusion could be established, several important consequences would follow. For example, the indiscriminate averaging of personal orders or rankings (the method adopted in so many studies of aesthetic preferences and the like) would become an "illegitimate and meaningless

procedure" ((14), p. 44). Moreover, the conclusion itself would seem, at least at first sight, to be quite incompatible with the main tenet of the Spearman school. On this ground, indeed, Stephenson and others originally rejected the method of correlating persons. He now, however, maintains that the factors revealed by the two methods "must belong to entirely different worlds", and supports his belief by a number of highly suggestive arguments. Nevertheless, before we decide to accept either of these rather drastic corollaries, the conclusion itself deserves a more searching scrutiny.

Curiously enough, no one has hitherto undertaken a comprehensive survey of the existing data in the way Spearman once surveyed all the available tables obtained by correlating tests. The object of this paper is to present the results of such a review, in the hope of reconciling antagonistic opinions on this and several minor but cognate controversies.[1]

II. Methods

To a large extent the differences of view seem attributable merely to differences in experimental or statistical procedure. In this country, for example, most investigators who have factorized correlations between persons have followed one or other of the multiple-factor methods originally outlined by Burt for that purpose (in his laboratory *Notes on Factor Analysis*, afterwards partly incorporated in (13), Memorandum I); and have used either his method of least squares (when dealing with covariances) or his summation formula (when dealing with correlations), seeking as close a fit as possible by means of successive approximation.

[1] It will be evident that I am deeply indebted to very many persons for their assistance. More particularly, I must acknowledge my gratitude to Dr Stephenson, formerly Research Assistant at University College. It was originally planned that my own investigation should be carried out under his immediate supervision, but his later duties at Oxford prevented any full or personal discussion; nevertheless, I am greatly in his debt for the trouble he has taken in sending detailed and stimulating criticisms of my results and for the kindly way he has always received incidental enquiries or criticisms of my own. He has also allowed me to make use of his own unpublished notes on *Q*-technique and of recent unpublished tables compiled by others working under his guidance (Miss Knowles, Mr Cohen, and Miss Hill). I am further indebted to Prof. Burt for allowing me to make full use of his notes on factor analysis and correlating persons, and for procuring material collected by his former students at the Institute of Education and elsewhere; to Prof. Valentine for putting me in touch with similar researches carried out at the University of Birmingham; to Prof. Beebe-Center and his statistical assistant at Harvard, Dr Harsh, for their helpful replies to my enquiries about work in America; and finally to many other investigators who have so willingly sent me tabulated data, whenever the figures in their publications or their theses were not sufficient for my purpose.

In America Beebe-Center at first[1] employed the earlier and simpler two-factor method devised by Spearman for analysing correlations between tests, and 'inverted' it for persons. Stephenson, in his first short paper on testing artistic taste, followed Burt and Bulley as regards both problems and methods: but in his longer article in this *Journal* (15), though the detailed formulae given are identical with those of Burt for a *single* factor, the proof by 'inversion' follows that of Beebe-Center.

His incidental criticisms of these two writers indicate the essential changes in his statistical approach. First of all, whereas Burt, Dewar, and even Beebe-Center (so I gather) in his later work, believe that an analysis by a multiple-factor technique is essential if additional factors are present, Stephenson maintains that "these generalized forms of factor analysis, such as that developed by Thurstone, have an artificiality which I cannot meet" ((15), p. 350). Secondly, whereas Burt maintains that, when we are correlating persons, we are in effect simply dealing with the 'transpose' of the initial matrix and therefore must reach practically the same factors as would be obtained by correlating traits, Stephenson now insists that the two systems "are the very poles apart". Burt's method, he asserts, like those of Spearman and of other psychometrists, can logically issue only in a complete denial of psychological types and type factors ((16), p. 295).

After reading my earlier draft of this section Dr Stephenson has been good enough to send a fuller explanation of his most recent views.[2] From his articles alone it is not always clear where he is merely simplifying Burt's original analysis for the aid of the reader unfamiliar with matrix algebra, and where he is definitely departing from it. Hence some of his pronouncements seemed at first sight a little inconsistent. (E.g. in his first paper on "Inverted Factor Technique" ((15), p. 345) he apparently follows Burt in assuming that the (normalized) factor matrix obtained on correlating persons may be regarded as an inverted form of the population matrix obtained by correlating tests (i.e. in the conventional symbols $L_p = P_t^{-1}$). But his reference to 'inverting' the initial matrix—and the theorems—is surely a misunderstanding. In substituting 'P-axes' for 'T-axes', Burt speaks of 'inverting' not the initial matrix (M_3), but the population and factor matrices M_1 and M_2 ((13), p. 253, footnote 1), i.e. $LV^{\frac{1}{2}}$ and P: no doubt the original choice of symbols was partly responsible for the confusion: in both 'P-technique' and 'Q-technique' we start not by 'inverting' but merely 'transposing' M_3.) Again, in his discussion of 'experiment 4' ((15), pp. 359–60; Cohen's investigation) he fully accepts the view expressed in Cohen's thesis that Q-technique will merely enable us to "find those persons who have led to the factors in R-technique and determine how many they are". And in his later article ((18), p. 195) he still speaks in the text of the 'second system' as being the "obverse or inversion" of the first; he also accepts Burt's proposal to standardize the matrices in terms of their 'unitary standard deviation' ($\sigma\sqrt{N}$) instead of the

[1] In his latest paper ((22), p. 353) he suggests substituting Thurstone's method of factor analysis; and, from a private communication kindly sent me by his statistical assistant, Dr Harsh, I gather that, in a series of experiments, not yet published, this method has indicated at least four factors operating in judgements on aesthetic value.

[2] This paragraph forms a later insertion, summarizing subsequent correspondence between Dr Stephenson and the author.

ordinary standard deviation (σ) ((13), p. 272)—a device which renders the matrix equations much simpler. But in the postscript (p. 206) he insists that "system 2 is *not* to be regarded as the 'obverse' or mere 'transpose' of data already analysed by system 1": and the preliminary standardization, he considers, must destroy any predictable relation between the two factor and population matrices deduced by the alternative approaches. In explanation he now points out that in the body of his article he was merely concerned to distinguish the main methods of standardizing and correlating the data in any matrix. This was already in type before Burt sent him his later manuscript(20), which was written just before his own postscript. The postscript, he explains, was thus added "in reply to the new arguments which seemed to destroy the uniqueness of the very method they were describing"; and is to be taken as his own "final view on this particular issue". I gather, too, that, although in his account of '*Q*-technique' he employs the matrix notation and (so far as possible) the symbols used by Burt and Thomson (in order to make the points of agreement and divergence clearer), he nevertheless has sometimes had to give the terms and symbols a special meaning of his own: thus, they must be "understood from their context rather than from a textbook of matrix algebra". He adds: "If no additional factors were brought to light by correlating persons, then person-factors can have no interest"; and the idea of deriving factors from "unstandardized scores and covariances" he rejects most emphatically as "scarcely meriting a moment's serious consideration". For this reason I have thought it necessary to base my own analyses throughout on correlations and not on covariances, even where the original investigators would have preferred the latter procedure. [In passing I must again express my indebtedness to Dr Stephenson for the fullness and readiness with which he has answered questions on these points.]

Similar divergences, not quite so clearly emphasized, are to be noted on comparing the procedures of other investigators; and must largely account for the conflicting results to which they have been led. But in addition to differences in method there seem also to be differences in the kind of test material chosen and the kinds of persons tested. Most investigators have worked (or assumed that they were working) with random samples of material and with random samples of the population; a few, however, have at times employed highly selected samples, in some cases exceptionally heterogeneous, in others exceptionally homogeneous. Now selection of itself can either eliminate 'factors' or introduce them. Here again, therefore, we have an unsuspected explanation for the contradictory conclusions. Stephenson, for instance, working in the same laboratory, used for his preliminary trials the pictures, cards, and lantern slides collected by Burt and Dewar (originally, I believe, in conjunction with Miss Bulley): but he was quickly led (so he explains) to "discard this material as unsuitable to my purpose" (the demonstration of group factors and mental types). The material substituted (e.g. simple colours, such as those used by Bullough in his studies on artistic types ((9), p. 280)) was usually far more homogeneous; and the altered figures show, what was after all to be expected, that this change was itself tantamount to partialling out one general factor. On the other hand, his groups of persons seem much more abruptly heterogeneous than those correlated by previous investigators. One group, for instance, includes five manic-depressive patients, five schizophrenic patients, and six normal members of a single family—two elderly and four relatively young ((17), p. 363); and this experiment, as we shall see, is one of the very few in which the specific 'type factors' are statistically significant.

The prior selection of types—of groups within groups, as it were—raises yet another issue. In analysing correlations between persons, as in analysing correlations between tests, the investigator has always to decide between two main alternative procedures—between what may be called the 'group-factor method' and the 'general-factor method' of analysis respectively ((13), p. 306). (*a*) With the first method, we partition the entire matrix of correlations into submatrices, and seek a separate factor for every submatrix that contains positive coefficients higher than those of the adjacent submatrices; this yields two or more distinct factors, each sharply confined to a certain group of persons, and possessing for those persons positive saturations only, and zero for the rest. (*b*) With the second method we analyse the matrix as a whole: this yields one or more general or 'universal' factors, shared by every person correlated. Whether tests or persons are correlated, most investigators have hitherto followed the second procedure (*b*). Burt apparently uses method (*a*) in correlating mental tests (e.g. (2), pp. 55–7) and method (*b*) in correlating persons, though in the earliest experiment on correlating persons (1), where marked visualizers, audiles and motiles were compared to demonstrate imaginal types, he apparently employed a group-factor method. Stephenson, on the other hand, invariably uses method (*b*) in deducing factors from tests, since "they are by definition universal" (i.e. general factors), but appears as a rule to adopt Burt's method (*a*) when correlating persons.

The grounds for the choice are seldom distinctly stated. Usually, it would seem, method (*a*) is adopted when there are negative coefficients in the correlation matrix. This, however, is not the true criterion. Contrary to a widespread notion, even when negative correlations appear, method (*b*) can in general still be applied; the result is then a bipolar factor. But, when we are correlating persons, the negative saturations that such a factor implies are no longer devoid of meaning, as they are generally supposed to be when correlating mental abilities. The proper criterion would seem to be rather this: method (*a*) is to be applied, when and only when there is definite evidence that the variables correlated form discontinuous groups—e.g. when the selected sample forms a highly heterogeneous population ((2), pp. 55–7, (24)). It has, however, recently been shown that, when the two alternative methods are applied to one and the same table of correlations, the two resulting factor patterns are related to one another in a very simple way, expressible by a triangular matrix; this implies that the positive and negative general factors really represent differences between each positive group factor and the remainder ("Factor analysis by submatrices", (24), p. 359).

In order that my own figures may be comparable throughout, I have used method (*b*) for every table, and method (*a*) in addition for those few cases where it seemed more especially appropriate. In calculating factor variances, I have used Burt's summation-method, inserting by successive approximation the full communalities. The results have been checked by proceeding to reiterated table-by-column multiplication, until as close a fit as possible has been attained. Since (as we shall see) most of the tables are nearly hierarchical, the variances for the first factor as thus computed here differ hardly at all from the figures that would be obtained by a simple summation method such as Thurstone's or Spearman's.[1]

[1] Where more than one factor is significant, the saturation coefficients thus found by the above method differ from Thurstone's chiefly in that the saturation coefficients are required to be uncorrelated. They are virtually those that would be reached by Hotelling's method or

To estimate the evidence for the *presence* of more than one factor, I have first examined the standard error of the residual correlations (or rather of their inverse hyperbolic tangents) according to Fisher's criterion for small samples ((11), p. 190). The residual is essentially a difference between two correlations—an observed and a theoretical. We therefore require the standard error of this difference. But the theoretical value is given by hypothesis, and so has no sampling error. Hence, if the traditional method is used, the residuals must be judged by the probable error of the observed correlation, not by that of the residual or of zero, as most investigators seem to have supposed. In Table I (p. 413) those investigations which yield significant residual correlations, on which a second factor can be based, are marked with an asterisk.

To estimate the *importance* of the various factors, the simplest method is to state the percentage of variance contributed by each. Although in most researches the evidence in favour of specific or group factors turns out to be 'non-significant', this does not prove (as is so often assumed) that such factors do not exist; the odds in favour of their existence nearly always remain higher than those in favour of their non-existence. The issue therefore becomes not: Is there or is there not a second factor? but: How large and how significant a part does it play as compared with the first or general factor? We may therefore agree with Stephenson in rejecting Beebe-Center's use of the tetrad-difference criterion and in seeking a 'more adequate method'. But precisely the same reasons compel us to reject the use of single-factor formulae (the so-called 'two-factor method') and to seek some means for estimating the relative importance of the multiple factors.

III. Data

In accordance with the foregoing principles I have made a complete factor analysis of all the tables of correlations between persons to which

by Kelley's, if *communalities* were substituted for reliabilities or unity. Except for the determination of appropriate communalities, the calculations are very much shorter than those described by Hotelling or Kelley. In computing the 'higher moments' ((21), pp. 185 *et seq.*), 'table-by-column multiplication', i.e. $R \times (Rt)$, can be substituted for 'table-by-table multiplication', i.e. $(R \times R)\,t$, unless the figures are to be taken to many decimal places. Where variances alone are needed, one may begin with table-by-column multiplication (as required for the shortened method of 'higher moments'), and then proceed to a column-by-column multiplication; thus, throughout Table I, I have taken

$$v = t'R^3 . R^3t / t'R^3 . Rt,$$

where t' as usual denotes the summation-operator [1, 1, ..., 1]. (The detailed working instructions are given in Burt's *Notes on Factor-Analysis*, which can be obtained from the Psychological Laboratory, University College.)

I could gain access. In all they number forty-eight. Spearman's first enquiry upon correlations between tests was carried out on fourteen (see (7), p. 138), and his second on twenty (see (7), p. 139), tables; so that it is fully time that a corresponding survey was carried out for correlations between persons. The tests surveyed by Spearman were almost exclusively tests of intellectual or cognitive abilities. Here, however, as will be seen from the headings in Table I (p. 413), the mental traits examined range over a much wider field.

The earliest work on correlations between persons was (as might be expected) concerned more with cognitive than with conative capacities. At the outset the method seems to have appealed mainly to educational psychologists interested in problems of group factors and mental types, and not wholly satisfied to ascribe all intellectual differences (except what was specific to each test) to a single factor, *g*. The earliest use of the method appears to be that of Burt and Davies in 1912 for a study of imagery types among school children ((1), p. 373); and the first table of this kind to be printed in full and examined from the standpoint of factor theory is that discussed by the former in his 1921 Report on *Mental and Scholastic Tests* and based on data obtained with tests of intelligence. An allied method was also used to obtain 'profiles' for various types of scholastic disability, based on ranking test items (or school subjects) in order of difficulty, performance, or preference, for individual children: in this earlier work, however, factor analysis, as now understood, was carried out only in exceptional cases. The majority of the earlier tables may be regarded as extended sets of consistency or 'reliability' coefficients, usually for intellectual abilities or attainments: in 1926, for example, Thomson & Bailes(6) published three tables of correlations between teachers' marks; though they do not actually work out saturation coefficients by the summation method, they give (according to a common practice at that time) correlations with the average or total order, which amounts to much the same thing. Some of the most interesting data are 'consistency coefficients' obtained from one and the same individual with one and the same set of test material. From 1922 onwards, first at the National Institute of Industrial Psychology, and later at University College, Burt and his research assistants, following a fairly systematic programme, have used the method for investigating temperamental types, aesthetic appreciation, and the like. Consequently, more recent researches, as will be seen from the list in Table I, deal rather with assessments of emotional tendencies or with subjective preferences—artistic, musical, literary, scholastic, and sensory.

The recent American work, carried out chiefly at Harvard, has also been concerned with 'preferential judgements' or 'likings', or with the reverse (i.e. 'dislikes', e.g. susceptibilities to different forms of annoyance: C. M. Harsh, unpublished communication). The 'preferential judgements' were concerned with preferences for odours(10) and, in later studies, for geometrical shapes, vases, poetry, and music(22). Similar work has been carried out by Pratt at Rutgers College and by Wilson at Western University, London (Ontario). I should add that the recently published experiments, carried out by Beebe-Center's students(22) and included in my table, were primarily designed to test Birkhoff's 'aesthetic measure' in various fields. Since the writers have explicitly refrained (p. 353) from applying any form of factor analysis (doubtless

because the number of items correlated was so few), it might seem unwarrantable on my part to factorize their tables and include the factor variances in my list below. In venturing to do so, my excuse is first that the results are strikingly similar to those obtained by other investigators who have applied factor analysis for themselves, and secondly that those who have professed to discover significant factors have often been content with correlating items that were just as few in number.

Stephenson's starting-point was Burt's 1935 Memorandum on factorizing evaluations of essays, etc., along the lines worked out for literary and aesthetic appreciation generally. He accepted the device of correlating persons (so-called *P*-technique) in principle, but strongly criticized the multiple factor technique, and proposed to substitute a modified technique of his own, which, he maintains, should lead to "an entirely new branch of psychometry, for which the term Personalistics may be coined". "There is nothing *statistically* novel", he writes, "about *Q*-technique, as there is in Burt's work on Person-factors; but the *psychological* implications of these factors have never been clearly appreciated, otherwise he would not have insisted on his reciprocity principle."[1] Accordingly, he himself decided to work over the same fields—aesthetic and educational preferences, emotional and clinical types—to demonstrate that the two sets of factors (the old *P*-factors and the new *Q*-factors) "have nothing whatever in common". The '*P*-technique' as used by previous workers erred by "still thinking in terms of the abilities or mental tendencies of *R*-technique": they consequently discovered only a few general factors instead of the innumerable type factors which *Q*-technique reveals. Hitherto one of the strongest supporters of Spearman's single factor theory, he now appears as one of the keenest defenders of multiple type factors. "Almost everything we factorists have denied about type psychology has been beside the point: types exist in great numbers" ((18), p. 209). He claims it as one of the chief merits of his modification that the investigations which, with the cumbersome method of multiple analysis, take the ordinary research student a couple of years can be "performed within an hour and factorized in two more". His own published work with the new method was carried out within a single year. As he himself points out, it should in consequence be regarded as consisting of "short extracts of experiments rather than completed studies"..."intended rather to illustrate a more adequate point of view than to draw final conclusions". So far, however, he himself has not yet extended his method to the cognitive field, with which his name has so long been connected—to the problems of verbal, perceptual, and perseverative types and factors, which he has hitherto investigated by *R*-technique (correlating tests). Nevertheless, his preliminary experiments, though brief, are sufficient, he believes, to reveal that "no one has previously used the technique of correlating persons either completely or with full understanding"; Thomson (in a recent paper criticizing Stephenson's rejection of group factors when correlating tests) had, he admits, "been very near to it, but saw only difficulties instead of a flood of immediate applications" ((15), p. 352).[2]

It is plain, therefore, that one of the primary objects of this paper must be to

[1] Memorandum on '*Q*-technique' (privately communicated: here and elsewhere I am indebted to Dr Stephenson for permission to quote from his unpublished notes and communications).

A fuller history of the method of correlating persons is rendered unnecessary by Vernon's thorough survey (recently published) of the various methods of assessing personal qualities ((30), esp. pp. 37–41, 63–5, 77: cf. also the historical introduction prefixed to (18)).

enquire how far the arguments and examples adduced by Stephenson do really, as he himself contends, at once "invalidate all previous efforts and overcome their difficulties". The fundamental issue that he raises might, I think, be stated as follows: When we correlate persons instead of tests, is the greater proportion of the factorizable variance still attributable to the first or 'general' factor (i.e. to one *g* and *n* specifics), or is it essential to invoke further factors in addition, each contributing a large enough proportion of the variance to be statistically significant?

IV. Results

The results of the analysis are summarized in Table I.[1] The answer to the main problem appears simple and quite definite. In the vast majority of cases, it will be seen, there is conclusive evidence for a single factor only—the first or 'general' factor. In five cases, however, there are significant indications of a second factor as well. In no case is there acceptable evidence in favour of a third. The average contributions to the total variance are, for the first factor, 0·48, for the second, 0·11, for the third, 0·04. Thus, as with correlations between tests, one 'general factor' almost invariably stands out as responsible for the majority of the variance; more specific factors, though doubtless present, are so uninfluential as to be exceedingly hard to detect.

In discussing certain conclusions, drawn by Mr Cohen ((26), cf. (15), p. 360) and myself, Dr Stephenson has argued that, even when the residuals are too small to be significant, nevertheless, if they fall into a definite bipolar pattern ('system 5' as it has been called ((15), p. 349)), then this pattern of itself demonstrates the presence of a bipolar factor, since such patterns could not be produced by chance. The argument,

[1] The correlation tables on which my figures are based, with full details as to saturation coefficients, etc., are given in the second part of my thesis(31). Tables from earlier enquiries have proved difficult to procure; hence researches prior to 1935 are, I fear, very inadequately represented. Owing to the expense of printing or typing the large correlation tables necessary when persons are correlated, most editors, it would seem, have been reluctant to print them in full, and many research students have not included them in their typed theses. I have, however, been able to procure manuscript figures for a few early investigations and a number of the most recent. The references at the end of this paper indicate the enquiries in connexion with which the data were collected. For the calculations I myself am primarily responsible in every case: indeed, comparatively few investigators have attempted factorization, and those who have done so have generally limited themselves to one factor or one small sample.

On the two particular problems that have given rise to the sharpest differences of opinion—olfactory preferences as studied by Beebe-Center and pictorial preferences as studied by Dewar—I have attempted to secure first-hand data of my own. The details are fully described in the first part of my thesis; here I need only say that in most essential points they seem to confirm the views expressed by those writers.

Table I. *Amount of variance attributable to first and second factors respectively in correlations between persons*

	First factor	Second factor
(i) Types of imagery: unselected children (Burt & Davies(1))	0·32	0·14
(ii) Types of imagery: selected cases (Burt & Davies(1))	0·28	0·23*
(iii) Difficulty of Binet tests: examiners (Burt(3))	0·88	0·04
(iv) Difficulty of Binet tests: children (Burt(3))	0·76	0·08
(v) Reliability of essay marks (*a*) (Thomson & Bailes(6))	0·63	0·02
(vi) Reliability of essay marks (*b*) (Thomson & Bailes(6))	0·64	0·06
(vii) Reliability of essay marks (*c*) (Thomson & Bailes(6))	0·51	0·07
(viii) Reliability of examination marks (Burt(13))	0·76	0·15
(ix) Reliability of examination marks (*a*) (Rhodes(13))	0·42	0·08
(x) Reliability of examination marks (*b*) (Rhodes(13))	0·28	0·13
(xi) Reliability of examination marks (*c*) (Rhodes(13))	0·49	0·15
(xii) Reliability of examination marks (*d*) (Rhodes(13))	0·48	0·13
(xiii) Tactile judgement (Binns & Burt(4))	0·43	0·05
(xiv) Preferences for pictures: children (Burt(9))	0·73	0·06
(xv) Preferences for picture postcards: students (Stephenson(15))	0·23	0·16
(xvi) Preferences for polygonal figures: laymen (Swift. Quoted in (22))	0·48	0·13
(xvii) Preferences for polygonal figures: students (Swift. Quoted in (22))	0·46	0·07
(xviii) Preferences for vases: students (Schnittkind. Quoted in (22))	0·33	0·15
(xix) Preferences for vases: naïve observers (Schnittkind. Quoted in (22))	0·37	0·15
(xx) Preferences for vases: students of fine art (Schnittkind. Quoted in (22))	0·51	0·08
(xxi) Preferences for pictures: art critics (Dewar(25))	0·70	0.04
(xxii) Preferences for landscapes: 1st Series: students (Eysenck & Davies, see (31))	0·31	0·10
(xxiii) Preferences for landscapes: 2nd Series: students (Eysenck & Davies, see (31))	0·22	0·11
(xxiv) Preferences for literary extracts (Williams *et al.*(27))	0·55	0·18
(xxv) Preferences for poetry (Miller. Quoted in (22))	0·34	0·13
(xxvi) Preferences for nonsense lines (Miller. Quoted in (22))	0·50	0·14
(xxvii) Preferences for musical extracts (Winter & Wright. Quoted in (27))	0·49	0·11
(xxviii) Preferences for musical extracts (Wing(33))	0·61	0·12
(xxix) Preferences for diatonic harmonies (Fischer. Quoted in (20))	0·37	0·15
(xxx) Preferences for humour (Schonfeld(35))	0·44	0·08
(xxxi) Preferences for smells (Beebe-Center(10))	0·57	0·08
(xxxii) Preferences for smells (Stephenson(15))	0·50	0·09
(xxxiii) Preferences for smells (Davies(31))	0·58	0·07
(xxxiv) Preferences for school subjects (Burt(8))	0·52	0·16
(xxxv) Preferences for school subjects (Pritchard(12))	0·47	0·11
(xxxvi) Preferences for school subjects (Stephenson(14))	0·45	0·18
(xxxvii) Tests of relation-finding (Cohen. Quoted in (15))	0·24	0·02
(xxxviii) Intellectual characteristics: normal adults (Burt & Spielman(5))	0·67	0·02
(xxxix) Intellectual characteristics: normal children (Burt(3))	0·51	0·08
(xl) Temperamental characteristics: normal adults (Burt & Spielman(5))	0·64	0·01
(xli) Temperamental characteristics: delinquents and neurotics (Burt(23))	0·60	0·29*
(xlii) Moods: normal adults (Stephenson(15))	0·46	0·09
(xliii) Moods: students (Knowles(32))	0·56	0·05
(xliv) Moods: normal children (Hill(34))	0·49	0·12
(xlv) Moods: psychotics and normals (Stephenson(17))	0·32	0·24*
(xlvi) Psychic tempo (Stephenson(16))	0·24	0·16
(xlvii) Moral reliability (Stephenson(17))	0·41	0·13
(xlviii) Children's fears (Campbell(28))	0·39	0·22*
Average	0·484	0·112

* The second factor is statistically significant.

however, appears to overlook the scheme of conditions that must inevitably be fulfilled if any set of intercorrelations, whether arising from chance or not, are to be consistent. These 'consistency conditions' from their very nature tend to produce the appearance of a bipolar hierarchy.[1]

Those, therefore, who have criticized previous investigators for confining themselves mainly to a first or 'general' factor and have maintained the overwhelming importance of the type factors seem so far to have failed in making out their case. To attempt a detailed examination of the several researches in turn would be impossible within the limits of a single paper. But a glance at one concrete problem, which has formed the centre of controversy in a series of three or four articles, will make the general position clearer. Working at Birmingham under Prof. Valentine's supervision, Pritchard(12) endeavoured to study pupils' preferences in regard to school subjects. The 8000 children were classified according to age, sex, and school, and the rankings averaged for the different age-groups somewhat on the lines of previous investigators(8). Although this procedure has hitherto been regularly employed by investigators who have dealt with this and similar problems, Stephenson's conclusions would render it extremely precarious. There must, he argues, be different 'types of preference' within each given age-group; hence it is not justifiable to average all the children of a given age without first applying '*Q*-technique' to discover the particular type to which each of the children belongs. Similarly, the attempts by Burt, Bulley, Dewar and others to measure capacity for artistic appreciation are "foredoomed to failure". Capacity for artistic appreciation is not a general factor; hence (as popular opinion rightly maintains) there can be no agreed or common standard. To illustrate these criticisms, he correlates the school preferences of fifteen boys and twenty-two girls, and discovers two distinguishable types, namely, the scientific and the linguistic(14).

Now, Pritchard and his predecessors have never denied the existence of such a bias in many individual pupils. Several have expressly noted it, and have pointed out that it forms the implicit basis of the familiar division into classical and modern sides and into Arts and Science Faculties. Thus, the essential question is, not whether such differences exist, but whether they are large enough, wide enough, and numerous enough to invalidate the general average. An answer can readily be obtained. If we take the tables printed to demonstrate the two contrasted types and partial out the general factor (which Pritchard's average is intended to measure), we then discover that not a single significant residual remains on which the two types can be convincingly based. Nor does the bias extend to more than two or three subjects, or affect those subjects very greatly. We are, indeed, given the order of preference for the children representing the two types; but the correlation between these orders proves to be nearly 0·5. The types, therefore, are not antithetical nor greatly inconsistent: in the 'scientific' type, physics, chemistry, and mathematics go up 2 or 3 places; French and English go down; but for the rest the general order is almost the same as in the 'literary' type. I may add that these conclusions are confirmed by the

[1] This is demonstrated at length in the appendix to my thesis(31). It is also very evident in Mr Cohen's first research (as cited in (15), p. 360), where only 12 items are correlated for 200 persons. Cf. also (22), p. 349, Table IX, where 88 out of the 105 coefficients fit a bipola pattern.

more recent work of Shakespeare (also from the Education Department at Birmingham) who writes that in his own data "the correlations (between persons) approximate to a chance distribution, and provide no significant divisions like Dr Stephenson's sharply divided groups in pupils of one school" ((19), p. 149).

Perhaps the acutest conflict, at any rate upon the surface, is between the views of Beebe-Center and Stephenson. Beebe-Center had adopted precisely the same inverted technique as that subsequently employed by Stephenson, and, indeed, had gone further, in that he applied Spearman's tetrad-difference criterion to check the results. Like Pritchard and others, he finds that individual preferences (this time for odours) are correlated highly enough to justify averaging, but discovers no evidence for more than a single factor. Stephenson has repeated Beebe-Center's experiment, and finds at least two factors, the second being relatively specific and bipolar. He claims that Beebe-Center, in describing the results in terms of Spearman's '*r*-technique', has overlooked the possibility of a multiplicity of factors, 'which looms large in *Q*-technique though rare in *r*-technique'. Unlike Beebe-Center, however, he applies no criterion to determine whether his second ('specific' or 'group') factor is significant or not, which is a little puzzling, since, when correlating tests instead of persons, he always rejects specific factors if they are not corroborated by the tetrad-difference criterion. In Table I, I give the variances for the two factors as obtained from Stephenson's own tables ((15), Tables II and III, pp. 356–8); and here again, after the main general factor has been eliminated, the residuals on which the types are based ('Factor II') prove to be all without statistical significance. It is only fair to add that Dr Stephenson's criticisms are reinforced by more general arguments ((15), pp. 353, 355); but, as I gather from Dr Harsh, these are partly due to an "excusable misunderstanding of Beebe-Center's viewpoint". The number of test stimuli on which the correlations were based is understated; and further it is wrong to suppose that Beebe-Center was seeking to discover "the persons with the most ability", i.e. "with the keenest olfactory susceptibilities"; his general factor was intended to measure (like Burt's and Pritchard's) not individual ability, but individual agreement with the general order. In a reply which he is good enough to forward on Prof. Beebe-Center's behalf, he describes more recent experiments with a Thurstone technique, and adds: "we certainly should not deny the existence of second factors or types... but are withholding judgement until we see if such factors operate in other groups."

This does not mean that Stephenson's very suggestive criticisms can be forthwith dismissed. The opposing views, I think, may easily be reconciled in the following way: we may agree with Stephenson that specific type factors probably exist; indeed, many of those that he mentions have already been indicated by correlations between tests (cf. (20), p. 65). At the same time we may agree with Beebe-Center and Pritchard in believing that, compared with the general factor, their influence is so slight as to be relatively insignificant in most of the experiments so far made. By a special choice of test material or of the individuals tested, we may, no doubt, bring them into prominence. But even so, like all such 'factors', they will be in the first instance mere statistical abstractions; they will be descriptive of the particular correlation matrix, but

not necessarily descriptive of the individuals or the traits. To analyse a table is not to analyse persons. We require first to prove that the classification thus suggested is a stable one, and secondly to demonstrate its psychological nature.

The stability of the types may be tested in the following way. Burt has shown that, if two correlation matrices A and B have exactly the same factors, then they must commute; i.e. their products AB and BA must be equal and therefore symmetrical ((20), p. 81). I have given one and the same set of students two different sets of pictures to rank in order of preference, and have correlated the results. The two correlation tables admirably fit the criterion; only one column of products shows deviations from the requirements of symmetry by more than three times the probable error, and that relates to the preferences of one particular person (myself) who deliberately changed the ground of choice in the second experiment.[1]

To determine the nature of the factors, the statistical study must be followed by, and if possible preceded by, a psychological study. Indeed, as a general rule, it is only when the numerical tables are supplemented by introspective data (as in (25)) or by an intensive clinical study (as in (23)) that the psychologist becomes convinced of the existence of the 'type tendencies', and begins to discern their concrete character. As Bartlett has observed: "The practice of separating observers into types may have a definite value, since it may help us to handle a mass of data easily. But we have no right to regard them as psychological starting points, unless we have better reasons for this than the mere classification itself.... Thus, a type having been distinguished, we still have to discuss the interrelation or characteristics by which it has been built up, or, failing that, give adequate reasons for regarding it as innate" ((29), p. 39).

From the data already accumulated, we can, I think, draw many instructive inferences, which may be helpful for further research. Briefly epitomized, the following are the main conclusions that emerge.[2] Most of them can be verified by the reader from a study of the figures given on p. 413, if he is familiar with the general problems and methods of the various investigations (see also Davies (31)).

[1] Detailed figures, together with a full description of the method, will be found in my thesis (31); for this part of the investigation I am deeply indebted to the assistance of Mr Eysenck.

[2] Those familiar with Burt's *Notes on Factor-Analysis* will recognize that many of my conclusions follow his, at times almost verbatim. But that seems unavoidable, since (as was to be expected) the more obvious and constant features such as were discernible in his briefer review would necessarily reappear in a wider one. I am specially grateful to him for permission to quote or modify his statements.

V. Summary of conclusions

A. *As regards methods*

(1) Where the persons correlated form a random sample, the method of factor analysis should proceed by an analysis of the correlation matrix *as a whole* ('method *b*'); but where the persons correlated are unavoidably compounded of samples drawn from distinct and discontinuous groups, and particularly where there is independent evidence showing the nature of this grouping, there a separate analysis of the positive submatrices ('method *a*') may prove the more effective procedure. With one or two exceptions, this double rule conforms to the practice of all the investigators in our list.

(2) In all the cases reviewed, the 'factors' which result from the statistical analysis appear to be capable of an intelligible psychological interpretation, *as they stand. Prima facie*, therefore, there seems to be no need for "rotation of the axes of reference": nor apparently has any investigator who has analysed correlations between persons attempted such rotation. (This, of course, does not preclude the possibility that rotation may be appropriate to future problems.)

(3) In two important respects the factor analysis of correlations between persons appears *simpler and more direct* than that of correlations between traits:

(*a*) The 'bipolar' factors (factors with negative as well as positive saturation coefficients) have an immediate and intelligible interpretation where correlations between persons are concerned, even if (as many have maintained) this is not so when correlating cognitive tests. Consequently, in correlating persons there is no need to "rotate the axes until there are no negative loadings".

(*b*) There is little reason to believe that any person should have a measurement of precisely zero for any particular factor. Consequently in correlating persons there is equally no need to "rotate the axes until a maximum number of the factor loadings are zero".

B. *As regards results*

(i) *First or general factors*

(1) Except when based on selected material, correlations between persons, like correlations between traits, tend almost universally to display some approximation towards a *hierarchical* arrangement, that is to say, towards an arrangement in which the relative influence of the first or 'general' factor is much greater than that of any other. Hence

the influence of all special (or group) factors becomes exceedingly difficult to demonstrate, and, unless the correlated items are exceptionally numerous, proves nearly always to be statistically insignificant. In forty-four out of the forty-eight researches here examined the first factor will account for all the variance and covariance within the limits imposed by the probable error.

(2) In correlations between persons as in correlations between traits, the saturation coefficients for the first factor are, as a rule, *positive* throughout: only in exceptional cases (usually arising out of homogeneity of persons or test material) are they partly positive and partly negative, i.e. *bipolar*.

(*a*) *Positive or unipolar factors.* (3) When positive, the first factor appears to be descriptive of a tendency *common to the total population* (or at any rate to the total sample) as distinct from more specific tendencies characteristic of particular individuals, types, or subgroups.

(4) In certain investigations this factor may represent nothing but an irrelevant or artificial influence—an intrusive *error or halo-effect*, which needs to be eliminated before the more specific factors can be brought to light. In correlating traits, 'halo-effects' (as is well known) may centre about certain persons; in correlating persons, similar 'halo-effects' evidently gather about certain traits. In both cases the appearance of a 'spurious' general factor is produced. Such a result is particularly noticeable when the examinees are required to assess *themselves*. Traits with attractive names may be ranked high by all; traits with repugnant names may be ranked low by all. Even when the judges are estimating *other persons*, they tend to rank high (or low) for each particular person a trait which ranks high (or low) for the whole human race.

(5) In other investigations this general population factor may be of interest and importance for itself.

(i) As a factor of *reliability*, it indicates that all the judges are tending to estimate or grade one and the same quality and to use one and the same set of criteria.

(ii) As a factor of *preference*, it tends to show that individual likes and dislikes (e.g. aesthetic, ethical, and even scholastic and sensory preferences) are to some extent general or 'objective' (in the sense that they are not solely based on individual peculiarities and so not purely subjective).

In these two cases the superior predominance of the first or general factor justifies the practice of taking an average (or a weighted average) of the individual judgements and regarding this as an 'ideal' measure of

the 'true' merits of the examinees or of the test objects. The presence of different mental 'types' does not invalidate this procedure (as has been supposed), because the type factor contributes so small a proportion of the total variance.

(6) In both cases *the first or general factor usually accounts for about half the total variance* (between 30 and 70 %). The remainder of the variance is due, not to any one predominant positive factor or "to an obverse or antithetical type tendency", but to a number of bipolar factors whose influence is diminishingly small.

(*b*) *Bipolar factors.* (7) A bipolar factor appears as the first factor only when the persons or the objects judged form a comparatively *homogeneous* group, e.g. when, in estimating emotional traits, all traits have the same average value, or when, in judgements of preference for pictures, etc., all the pictures are of approximately the same general artistic merit. The selection of a group of items that are homogeneous in regard to the qualities to be judged is in fact equivalent to eliminating by experimental rather than statistical methods the general or positive factor, and so converting what would otherwise be a second factor into the first. Certain recent experiments suggest that this technical device supplies a useful method of procuring significant evidence for the bipolar factor, when the ordinary method (which leaves it to appear as a second factor rather than the first) is likely to fail.

(8) Under these conditions negative as well as positive correlations then appear in the original table. Such a pattern, however, is *exceptional*, and does not (as has been stated) form the general rule. And even so the table still approximates to hierarchical form. It is by no means necessary (as has been erroneously maintained) to introduce a separate specific factor for each positive matrix or for each separate type; if this is done, then the two factors, so far from being independent, prove to be highly correlated.

(9) In virtue of the 'consistency conditions' obtaining among correlations, the appearance of a bipolar hierarchy may arise among correlations whose varying values are *due exclusively to chance.* By itself, therefore, the appearance of such a pattern provides no evidence for the existence of a significant bipolar factor or of antithetical types.

(ii) *Secondary or specific factors*

(1) All factors except the first prove to be bipolar. When genuinely significant such bipolar factors may be taken as indicating the presence of two *antithetical types or tendencies.*

(2) Since in most published investigations an overwhelming proportion of the variance proves to be attributable to the first factor, it is not surprising to find that the saturation coefficients for secondary or specific factors generally have *little or no statistical significance*. The practical corollary seems obvious: when the method of correlating persons is used to discover so-called mental types, then either the number of traits must be very large, or the persons (and in certain cases the tests) must form a homogeneous group. When this is not feasible, then the types and the factors must be elicited, not by correlating persons, but by correlating traits.

(3) In aesthetic preferences, in children's preferences for school subjects, and in temperamental traits, different investigators seem to have reached quite independently much *the same factors or types*. Nevertheless, with rare exceptions, it would seem that their conclusions really rest more upon general observation than upon the statistical evidence which they adduce. The statistical evidence is consistent with their views, but rarely affords any proof.

(4) Where the same group of persons has been tested with different versions of the material and so has supplied two or more matrices of correlations, the symmetry criterion may be used to demonstrate that *the same factors are present in the different matrices*. It appears that the factors reached by correlating persons are by no means necessarily limited by the 'set up' of the test-material on which they are based, but on the contrary are comparatively stable for any given set of persons. Under certain conditions (viz. similar age, attitude, intelligence and variability of the different groups tested) it seems possible to show that they are also stable for one and the same set of tests.

REFERENCES

(1) Burt, C. (1912). *J. exp. Ped.* I, 355–88.
(2) —— (1917). *Distribution and Relations of Educational Abilities.* London: P. S. King and Son.
(3) —— (1921). *Mental and Scholastic Tests.* London: P. S. King and Son.
(4) Binns, H. & Burt, C. (1922). *J. Nat. Inst. Industr. Psychol.* I, 93–8.
(5) Burt, C., Gaw, F. & Spielman, W. (1926). *A Study in Vocational Guidance.* London: H.M. Stationery Office.
(6) Thomson, G. & Bailes, H. (1926). *Forum of Education,* IV, 85–96.
(7) Spearman, C. (1927). *Abilities of Man.* London: Macmillan and Co.
(8) Burt, C. (1931). Board of Education, *Report of the Consultative Committee on the Primary School.* London: H.M. Stationery Office.
(9) —— (1933). *The Psychology of Art,* ap. Burt, C., Jones, E., Miller, E., Moodie, W. *How the Mind Works.* London: Allen and Unwin.

(10) BEEBE-CENTER, J. G. (1933). *Pleasantness and Unpleasantness.* New York: Century Press.
(11) FISHER, R. A. (1934). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.
(12) PRITCHARD, R. A. (1935). *Brit. J. educ. Psychol.* v, 157–79, 229–41.
(13) HARTOG, P., RHODES, E. C. & BURT, C. (1936). *Marks of Examiners.* London: Macmillan and Co.
(14) STEPHENSON, W. (1936). *Brit. J. educ. Psychol.* vi, 43–57.
(15) —— (1936). This *Journal,* xxvi, 344–61.
(16) —— (1936). *Char. and Pers.* iv, 294–304.
(17) —— (1936). *J. educ. Psychol.* xxvii, 353–67.
(18) —— (1936). *Psychometrika,* i, 195–209.
(19) SHAKESPEARE, J. J. (1936). *Brit. J. educ. Psychol.* vi, 147–64.
(20) BURT, C. (1937). This *Journal,* xxviii, 59–96.
(21) —— (1937). *Brit. J. educ. Psychol.* vii, 172–95.
(22) BEEBE-CENTER, J. G. & PRATT, C. C. (1937). *J. gen. Psychol.* xvii, 339–54.
(23) BURT, C. (1938). *Brit. J. med. Psychol.* xvii, 158–88.
(24) —— (1938). *J. Psychol.* vi, 339–75.
(25) DEWAR, H. (1938). *Brit. J. educ. Psychol.* viii, 29–49.
(26) COHEN, J. I. (1938). *J. ment. Sci.* lxxxiv, 495–512.
(27) WILLIAMS, E., WINTER, L. & WOODS, J. M. (1938). *Brit. J. educ. Psychol.* viii, 231–49.
(28) CAMPBELL, H. F. (1938). *Brit. J. educ. Psychol.* viii, 314.
(29) BARTLETT, F. C. (1938). *Remembering.* Cambridge University Press.
(30) VERNON, P. E. (1938). *Psychological Tests of Personal Qualities.* London: H.M. Stationery Office.
(31) DAVIES, M. (1938). "The General Factor in Correlations between Persons." Unpublished Thesis, University of London Library.
(32) KNOWLES, E. (1938). "A Study of Extraversion and Introversion by *Q*-technique." Unpublished Thesis.
(33) WING, H. D. (1938). "Tests of Musical Ability." Unpublished Thesis.
(34) HILL, M. E. (1938). "A Genetic Study of Temperament." Unpublished Thesis.
(35) SCHONFELD, A. (1938). Unpublished work, University College, London.

(*Manuscript received* 15 *June* 1938)

TEMPERAMENT AND THE THRESHOLD FOR FLICKER

By LL. WYNN JONES

I. *The flicker test and mania-melancholia* (p. 422).
II. *The flicker test and introversion-extroversion* (p. 423).
III. *The flicker test and perseveration-non-perseveration* (pp. 423–424).
IV. *The flicker test and ideo-motor tests* (p. 425).
V. *The need of a more comprehensive investigation* (pp. 425–426).
References (p. 426).

The flicker test has been used in the study of melancholia, introversion, and perseveration, but strict definitions of these three terms and their antonyms are not available. Positive statements concerning them or their interrelations must therefore be formulated with care. It will be shown later that lack of such care easily leads to semantic confusion and to gross violations of the rules of Aristotelian logic.

I. The flicker test and mania-melancholia

As regards melancholia, let us assume that the staff of a large mental hospital were suddenly replaced by a new staff who, without access to the case records, had to label each patient anew. Would the resulting 'melancholia' group coincide with the original, even if typically circular cases were excluded from consideration? Let psychiatrists give the answer. "Psychiatric syndromes are so protean in their manifestations as to be differently classified by competent observers with only slightly different orientations" (1). "We are far from satisfied with our diagnoses" (2). "General formulae are dignified, and diagnostic terms give comfort, but they are verbal symbols which are apt to do violence to the complexity of the facts" (3). Such terms, however, are obviously essential to the psychiatrist. It is the ordinary psychologist who needs to exercise care in their use.

The flicker test was employed by Wiersma (4,5), who found the speed of fusion of colours to be much lower for melancholic than for manic subjects. The present writer (6,7), on the other hand, found no such distinction. It is true that Wiersma included cases of paranoia in his melancholic group, but this is not likely to be the sole cause of the discrepancy between the two findings.

II. The flicker test and introversion-extroversion

Washburn *et al.* (8) found that the speed of fusion was actually lower for extroverts than for introverts, but it must be remembered that introversion tests can hardly be regarded as yielding explicit measurements. Guilford (9,10), for example, found no single factor running through all the items of an introversion test.

III. The flicker test and perseveration-non-perseveration

As regards perseveration, the need of definition is also manifest. A paper by Stephenson (11) specially emphasizes this need. The work of Lankes (12) has served as a beacon light to all subsequent workers in this field. Lankes studied the interrelations of several phenomena and produced evidence that the speed of fusion was related to perseveration.

Quite recently Biesheuvel (13), by means of an ingenious device similar in principle to the stroboscope, also concludes that the speed of fusion yields a measure of perseveration. The present writer has no grounds for doubting the truth of this conclusion. Nevertheless, the aim of this communication is to consider some misconceptions which may obscure the issues when the flicker test is studied in relation to orectic traits.

Biesheuvel devotes more than a tenth of his article to a consideration of Lankes's method and makes the astonishing statement that "Lankes's procedure has been described at length because it was also adopted by Wynn Jones"! It is true that "I adopted several improvements due to Lankes such as descending from a high speed instead of ascending from a low speed", and I still consider that such a procedure yields reliable measurements in the hands of those acquainted with psychophysical methods. Biesheuvel has apparently failed to realize how widely my aims and procedures differed from those of Dr Lankes.

(1) My subjects were patients at a mental hospital. It was therefore necessary to use a simpler technique than that which served Lankes so well with students. An experienced tester like Lankes would not have considered a session lasting forty minutes when testing abnormal subjects.

(2) My subjects were tested individually.

(3) My disks were driven by an electric motor.

(4) I was concerned in comparing the reactions of manic and melancholic subjects in the flicker test, irrespective of whether it was a measure of perseveration. Such being the differences in procedure it is unnecessary for my present purpose to give further consideration to the criticisms of the Lankes technique, although some of them could be refuted. On

reading further I find that "Wynn Jones's contradictory evidence must be rejected, his apparatus and technique not being suitable for the measurement of the threshold for flicker". Now Biesheuvel's efforts to develop "a careful and standard technique" of using the flicker test will be commended by everyone. But he has obviously not realized what my evidence was intended to convey. Unlike Wiersma I found that the flicker test did not distinguish between cases of melancholia and cases of mania. I said nothing to indicate that it was not a test of perseveration. I therefore suggest to Biesheuvel a method of checking my conclusions. Let him try his own method, unless it is too complicated, with melancholic and manic subjects. But the following rules of the game must be observed. He must take a melancholic and a manic group duly selected by hospital doctors and no case must be rejected without reasonable cause. He should preferably not be told to which group a patient belongs until his fusion speed has been determined. If he found a result different from mine he would have some grounds for rejecting my evidence.

Biesheuvel also raises the important question of dark adaptation, but I failed to see any reference to the period which he adopted. Ross(14), in his well-controlled investigation at Yale, kept his subjects for two minutes in a lighted room; then there followed one minute of darkness before the stimulus was lighted. Although I tested each subject with red-green and with blue-yellow disks on five different days I cannot claim to have rigorously controlled the factor of adaptation. I was fully cognizant of its importance, for previous to that time I had probably had more dark-adapted subjects than most psychologists and had on many occasions acted as subject, e.g. as one of the three subjects of Dr Goldschmidt in his *Eigenlicht* researches in Wundt's laboratory. I now turn to the statistical significance of Biesheuvel's results. One weakness he has himself pointed out, namely, that the estimates of perseveration were supplied by only one judge. Moreover, the small sizes of his two groups preclude the drawing of a positive conclusion. His first group numbered only nineteen cases. We are not told how many of these were perseverators and how many non-perseverators. We are told that the difference between the average flicker threshold for the perseverators and non-perseverators was three times the standard deviation. Actually I find it was 2·6. The corresponding figure for his second group of twenty-three cases is given as four. Actually 3·9.

In the absence of any data on the intra-group variability, would any statistically minded investigator attach value to such a test for measuring perseveration in the *individual*? or even for measuring the difference in perseveration between two groups?

IV. The flicker test and ideo-motor tests

In advocating the claims of his test Biesheuvel further writes: "The introverted individual will be much more at ease observing and reporting a subjective experience than performing a task." He is thus contrasting the experience of judging flicker with performance in a motor test of perseveration. He seems unaware of the fact that Pinard, Cattell, and others have produced evidence that it is entirely wrong to connect high and low perseveration with introversion-extroversion (15,16). For Biesheuvel, melancholia, introversion and perseveration are apparently almost synonymous terms. I agree with him that the flicker test may be a good instance of Spearman's law, but it does not necessarily follow "that the threshold for flicker should provide a better measure of perseveration than the usual ideo-motor tests". This can only be settled by experiment. It is at least significant that since 1914, when I employed motor tests (17) embodying the principle of interference with a well-formed habit (e.g. the reverse stroke test and the "it" test), and the alternation principle (e.g. the 2S test), dozens of investigators have followed suit and have improved the technique. The reason for the prevalence of such tests is that they produce results and, as they are group tests, they produce quick results. One may, however, hazard the opinion that some of them would prove still more serviceable when applied to one subject at a time as is necessarily the case with the flicker test. There are so many overt movements in the motor tests which cannot be controlled when employed with groups. I am at present attempting to use tests when such movements are kept under control.

V. The need of a more comprehensive investigation

In conclusion I may say that it is no secret to many that I started a series of individual tests of various kinds about the same time as the group tests mentioned in the *Abilities of Man*. Amongst them were tests of the following phenomena: cold, heat, pain and pressure after-effects, the after-effects of seen movements, after-images and induction effects, e.g. when one tapping rate is followed by a faster. Unfortunately the work was interrupted before a sufficient number of subjects had been tested. It may be of interest to future investigators if I refer very tentatively to a few 'pointers'. There were two kinds of tests which curiously enough did not seem to correlate with the others, viz. tests of the after-effects of seen movements and induction tests.

It was also found that some of the after-effects studied required from the subjects a considerable degree of skill in introspection.

The results of this preliminary testing certainly indicate the need of a more extensive investigation into the interrelations of such phenomena and their bearing on Spearman's law of inertia: *Cognitive processes always both begin and cease more gradually than their (apparent) causes.* But there still remains the formidable but necessary task which has been recommended by Prof. Spearman, namely, to investigate further the scope of this law with reference to various processes or tasks.

REFERENCES

(1) Masserman, J. H. & Carmichael, H. T. (1938). "Diagnosis and prognosis in psychiatry." *J. Ment. Sci.* LXXXIV, no. 353.
(2) Bond, E. D. & Braceland, F. V. (1937). "Prognosis in mental disease." *Amer. J. Psychiat.* XCIV, 263.
(3) Campbell, C. Macfie (1935). *Destiny and Disease in Mental Disorders, with Special Reference to Schizophrenic Psychoses.* New York: W. W. Norton and Co.
(4) Wiersma, E. (1906). "Die sekundäre Funktion bei Psychosen." *J. Psychol. Neurol., Lpz.*, VIII, 1.
(5) —— (1932). *Lectures on Psychiatry.* London: H. K. Lewis.
(6) Jones, Ll. Wynn (1928). "An investigation into the significance of perseveration." *J. ment. Sci.* LXXIV, 653.
(7) —— (1929). "Individual differences in mental inertia." *J. Nat. Inst. Industr. Psychol.* IV, 282.
(8) Washburn, M. F., Hughes, E., Stewart, C. & Sligh, G. (1930). "Reaction time, flicker and affective sensitiveness as tests of extraversion and introversion." *Amer. J. Psychol.* XLII, 412.
(9) Guilford, J. P. (1933). "An examination of a typical test of introversion-extraversion by means of the method of similar reactions." *J. Soc. Psychol.* IV.
(10) Guilford, J. P. & Guilford, R. B. (1934). "An analysis of the factors in a typical test of introversion-extraversion." *J. Abn. Soc. Psychol.* XXVIII, 377.
(11) Stephenson, W. (1934). "An introduction to so-called motor perseveration tests." *Brit. J. educ. Psychol.* IV, 186.
(12) Lankes, W. (1915). "Perseveration." *Brit. J. Psychol.* VII, 387.
(13) Biesheuvel, S. (1938). "The measurement of the threshold for flicker and its value as a perseveration test." *Brit. J. Psychol.* XXIX, 27.
(14) Ross, R. T. (1936). "The fusion frequency in different areas of the visual field." *J. Gen. Psychol.* XV, 133, 161.
(15) Pinard, J. W. (1932). "Test of perseveration." *Brit. J. Psychol.* XXIII, 114.
(16) Cattell, R. B. (1933). "Temperament tests." *Brit. J. Psychol.* XXIV, 20.
(17) Jones, Ll. Wynn (1915). "Perseveration." *Rep. Brit. Ass.* p. 698.

(*Manuscript received* 8 *December* 1938)

A FURTHER NOTE ON THE DEGREE OF VARIATION IN THE SCORE OF A MOTOR TEST OF PERSEVERATION

By JANE DARROCH

In a discussion with the writer of her previous paper, "An Investigation into the Degree of Variation in the Score of a Motor Perseveration Test" (this *Journal* (1938), XXVIII, 248), Mr Babington Smith pointed out that the writer in her previous article, in order to estimate whether the score for the *w* test was sufficiently constant for the test to be of value, used as a criterion that the scores of high and low perseverators should not overlap with those of medium ones, and that this criterion was a very exacting one, and at the same time likely to be inexact, since the extreme cases of distributions tend to be unreliable, and that it was inefficient, since only a small part of the information collected was used. Mr Babington Smith proved, by the use of Fisher's analysis of variance, that the differences between the average scores for all the trials included in the tables in the previous article, that is, for about thirty trials, were greater than was due to chance; and pointed out that, since that was so, some importance could safely be attached to the differences between scores. He raised the question whether the same criterion was satisfied by the results of the first five, or any five consecutive, tests taken by the subject, this being a more manageable number of tests to give.

The writer accordingly applied the criterion of Fisher's analysis of variance to the average score for the first five trials, scored by the methods $(X_1+X_2)/Y$ and X_1/X_2, and to that for the trials from the sixteenth to the twentieth inclusive, scored by the same two methods. The score $X_1/X_2+2X_1/Y$ was ignored, because it is so like X_1/X_2 both in variability and in the rank order in which it places the subjects, that anything true of X_1/X_2 may be assumed to be true of it too. Only those eight subjects were included in the calculation who had had no previous practice at the test. For this number of subjects and of trials, there is a 1 % probability that Z exceeds 0·5481 as a result of pure chance. It was found that the mean scores for the first five trials did differ more than was due to chance when they were scored by the method X_1/X_2, Z being 0·7503; and that the same was true of the trials from the sixteenth to the twentieth inclusive, whichever method of scoring was used, since Z was 0·6506 for

$(X_1+X_2)/Y$, and 1·0614 for X_1/X_2. The average score for the first five trials did not satisfy the criterion when the method $(X_1+X_2)/Y$ was used, Z being in this case 0·4362.

It thus appears that, at least when the method of scoring X_1/X_2 is used, the differences between mean scores for five successive trials is sufficiently great for some reliance to be placed on the results of the test. As every trial was included in the calculation, whether any disturbing influence was present or not, except one trial which was mistimed, it also appears that disturbance by such factors as noise and fatigue can be safely ignored unless it is very gross.

As already noted, the rank order of the subjects changes considerably between the earliest trials and the later ones. For the score X_1/X_2, it was F, A, N, B, E, H, M, I for the mean score of the first five trials, and A, F, M, H, E, N, B, I for the mean score of the trials from the sixteenth to the twentieth. Thus, although each of these scores differentiates the subjects satisfactorily, they do not measure precisely the same qualities. The question remains: What is the cause of the difference, and which score is the more likely to yield a measure of some temperamental quality? It is noteworthy that all the statistically significant correlations so far obtained between perseveration tests, scored by the ratio method, have been between first and second trials.

PUBLICATIONS RECENTLY RECEIVED

Experimental Psychology. By ROBERT S. WOODWORTH. New York: Henry Holt and Co. 1938. Pp. xi+889. $3.80.

There is at the present day a widespread belief that experimental psychology has failed to fulfil its early promise as a scientific discipline. An alarming sense of disillusion is abroad, penetrating even into our laboratories. Our students acquire an ever-growing distrust in the confident enthusiasm of the early pioneers. Our 'qualified psychologists' only too often qualify their psychology with a tolerant resignation to its manifest shortcomings. Has, then, the mountain of Leipzig in labour brought forth but a mouse—or at best, a humble white rat?

Prof. Woodworth's recent book, despite its many and obvious qualities, will do nothing to combat this prevalent attitude of disillusion. Every psychologist feels intense gratitude to the founders of his science—to Fechner, to Wundt, to Ebbinghaus, for their untiring efforts to establish an experimental science of mind. But many of us find our admiration of these men not untinged by envy. Cannot we too develop a like spirit of optimism and confidence—if not in ourselves, then at least in our students? Prof. Woodworth's book does not help us here. It is no monument to the passionate belief held out by the pioneers of experimental psychology in the future of their science. Far rather is it a mausoleum of their methods. The author is dominated by the *means* devised by the pioneer experimentalist. What we require to-day is a reaffirmation of his aims and ideals.

Experimental Psychology opens with a chapter on Memory. This contains little else than classification of the 'exact methods' devised by Ebbinghaus and elaborated at great length by his followers. The next two chapters, on Retention and Memory for Form respectively, are replete with facts, both quantitative and qualitative. But the underlying theory is of the sketchiest. Woodworth's assumptions are the familiar postulates of every trace theory, and they are couched in the vaguest terminology of function. Indeed Woodworth's concept of the memory trace is about as 'precise and scientific' as is Ebbinghaus's technique of the nonsense syllable. It would be unfair to allow dissatisfaction with theory to blind one to the obvious merits of these three chapters on memory. There is an immense amount of material here, and the treatment is expressly adapted to the needs of the student who must perforce familiarize himself with the very great mass of controlled experimental data. The presentation, moreover, is both clear and intelligent throughout.

Chapter v deals with the Conditioned Response. This is far more constructive than any of the preceding chapters. The author has made a real effort to grasp, and clearly to set forth, the implications of conditioning for psychology as a whole. The latter half of the chapter comprises, indeed, one of the best general discussions which we have as yet had of the relation of the conditioned response to other forms of behaviour, and to the theory of learning in general. It is marred only by a rather crude and anthropomorphic 'sign theory' of the conditioned stimulus. Chapter VI, on Maze Learning, is also interesting. Prof. Woodworth takes, very properly, as his problem *how the maze is learnt.* His emphasis throughout is upon accomplishment rather than explanation, and the empirical treatment adopted is highly successful. Moreover, Prof. Woodworth's presentation of the experimental studies in historical sequence is of enormous assistance in enabling the reader to gain some sort of an orientation in a field where so much of the contemporary discussion is on the level of highly specialized and stylized theory. And the value of the chapter for the non-specialist reader is further enhanced by the excellent comparative treatment of maze-learning in animals and men. Chapter VII, on Practice and Skill, is competent though not very stimulating. A good deal of the standard objective data is summarized, but there is practically no discussion of the really vital theoretical issues underlying it.

There is likewise no mention of the intensely significant work of Michotte and his pupils on the 'morphology of movements'. This is a notable omission. Chapter VIII and IX deal with the Transfer of Training and Economy and Interference in Learning. There is an enormous amount of experimental material in these chapters, but one cannot avoid the suspicion that the problems are not really worthy of so detailed a treatment. The same suspicion attaches in even more marked degree to the immediately succeeding chapters on some aspects of feeling and emotion.

Chapter XIV, on Reaction Time, is of interest primarily from the standpoint of psychological history. Problems of reaction time are scarcely of central importance to-day, but one may venture to suggest that at no time will the psychologist be in a position wholly to dispense with these measurements. Prof. Woodworth's chapter deserves to become the standard reference for all who have to make use of reaction times in an experimental or clinical capacity.

Chapter XV is called (rather misleadingly) Association. We may hasten to point out that it contains no discussion whatsoever of the celebrated principles. The chapter is entirely devoted to systematizing the results of innumerable experimental studies of word-association. Its value as a summary is considerable, and this chapter, also, is doubtless destined for the fate of a standard reference.

Chapters XVI–XVIII are concerned with Psycho-physics. Here the author's obsession with the methods of Fechner and Wundt emerges most clearly. These methods may, it is true, have the 'disciplinary value' grudgingly accorded them by James, but what is required far more urgently at the present day is a redefinition of the *problems* of psycho-physics. Prof. Woodworth's discussion of psychological measurement, for all its superficial plausibility, manages to evade every serious criticism of the traditional arguments. And he is quite certainly insufficiently critical of Fechner on the psychological side. These chapters, admirable as in many respects they are, clearly point to the past rather than to the future. The present-day student may derive considerable instruction in method from a careful study of this section of the book. But for his inspiration he must needs seek elsewhere.

Chapters XIX–XXII deal with the Senses. The first of these, on the skin senses, is careful and detailed enough, though the recent and highly significant researches of Woollard are entirely omitted. Chapter XX, on Smell and Taste, is neither as complete nor as up to date as one might wish, and no recent work on threshold determination is quoted for either sense. Chapter XXI, on Hearing, is a brave effort to do justice to a field which is daily becoming more complex and specialized. The expert will note some inaccuracies of fact, and he may well be critical of Woodworth's theoretical interpretations. European workers, moreover, may resent the somewhat scanty treatment accorded to their contributions. It is further to be regretted that Prof. Woodworth has not seen fit to include any discussion of the implications of *Gestalt* psychology for auditory theory. Chapter XXII, on Vision, gives a fairly accurate general discussion of the psychological phenomena. It is, however, somewhat disjointed in tone, and the reader must remain content with a rather superficial evaluation of contemporary theory. The author indeed appears to favour the Ladd-Franklin theory of colour vision, despite the weight of evidence against it. Eye Movements are treated with great thoroughness in Chapter XXIII, though the space devoted to this topic may seem to some excessive. Some problems of visual perception are dealt with in Chapters XXIV–XXVI. In the first of these, the author discusses the appreciation of colour and brightness. There is a very good summary of the principal experiments on brightness constancy, and the value of the chapter as a whole is enormously enhanced by the numerous and excellent diagrams of laboratory set-ups. Theory as usual takes second place, though it might seem to some of us that Koffka's ingenious speculations deserve a more thorough and detailed examination than they actually receive. In the following chapter, the organization of two-dimensional fields receives some consideration. This is a highly eclectic chapter, but it has value as an impartial summary of researches on figure and ground, contour and shape, and the conditions of 'organization' in general. Six theories of perceptual organization are listed at the outset, but little attempt is made in the remainder of the chapter to evaluate these theories in terms of the available data. Nevertheless, the chapter as a whole should be of great value to the student,

and the exposition is clear and straightforward throughout. Chapter XXVI, on Visual Space, is notable for an exceedingly able presentation of the principal conditions making for tri-dimensional experience. In a field such as this, probably the most that the writer of a text-book can do is to set out the data clearly and to assess as far as is possible the relative parts played by the various visual and non-visual 'cues' in the perception of distance. And in such a task as this Prof. Woodworth excels.

Chapter XXVII is concerned with Attention. There is much to be said for the resurrection of this concept in a modern text-book. At last psychologists are beginning to realize the absurdity of a psychology in which the term 'attention' is exclusively used in the instructions given to their subjects. But there is disappointingly little that is new in Woodworth's treatment. Our old friends—span of apprehension, shifts and fluctuations, 'doing two things at once'—feature once again. But that is all. If the notion of 'attention' is to reinstate itself honourably in psychological discussion, it must hold forth promise of greater fruitfulness than this.

Chapters XXVIII–XXX deal with Reading, Problem Solving Behaviour, and Thinking respectively. The last two chapters are quite adequately done, though a specific lack would seem to us the absence of any account of the relations of language and thought, and of language itself. But these two chapters do bring together a mass of experimental data bearing upon the thought processes, and taken together, they furnish an excellent introduction to the experimental study of the higher mental processes in general.

Experimental Psychology, despite its author's long-standing and distinguished connexion with American 'Functionalism', is not written primarily from the standpoint of biological science. The heritage of Fechner and Titchener, of the physicist and the formalist, lies heavy upon almost every page of this long and significant book. The really vital problems of present-day psychology—the problem of levels of functional integration, the problem of cortical function in learning, the problem of 'sensory transfer' and the generalization of response, the problem of 'schemata' and the co-ordination of reaction, and, most pressing of all, the problem of motivation in the widest sense—these find no place at all in Woodworth's text. Even the one outstanding theoretical movement of the present century, the *Gestalt* psychology, is treated throughout more as a collection of unrelated suggestions than as a well-organized body of fact and theory. We may in all sincerity congratulate Prof. Woodworth on his diligence, his competence and his erudition. But we find it hard to sympathize with his choice of the material in a book which must surely become for many years the senior text-book in the English language.

Prof. Woodworth is a brave man. And he deserves the profound gratitude both of teachers and of students in psychology. His significant book will long be remembered, if only as the epitome of a phase in which psychology passed through a period of arrested development—a kind of 'plateau' in its progress. And when all is said and done, Prof. Woodworth has achieved more than any other single man since Hermann Ebbinghaus. *Experimental Psychology* is destined for a place on the shelf beside the *Elemente*, the *Grundzüge*, and the *Principles*. O. L. Z.

General Psychology. By W. J. H. SPROTT. London: Longmans, Green and Co., Ltd. 1937. Pp. vii + 446. 7*s*. 6*d*. net.

Psychology for Everyone. By W. J. H. SPROTT. London: Longmans, Green and Co., Ltd. 1937. Pp. xi + 446. 8*s*. 6*d*. net.

These two books are identical except for their prefaces. The first is addressed to the student *and* the general public, and the second to the general public alone; the student therefore has the privilege of buying almost the same material for a shilling less. It is to be hoped that this privilege will be very widely exercised, for as a starting text-book in the subject this is a very good and comprehensive volume. It demands, indeed, the exercise of considerable intelligence on the part of the reader. Mr Sprott does not shrink from controversial issues, he does not disguise them, and he presents many

sides to most problems, often in an exceedingly concise manner. Since, however, nobody is likely to make any headway in psychology unless he is forced to think for himself, all this is to the good.

The book has an original and rather provocative beginning. Everybody knows that psychologists are constantly taunted with persistent internal disagreements. Mr Sprott wisely faces this straight away. He asserts that the differences between 'schools' of psychologists are often much less important than the differences of label would suggest. He states what the main schools are, proclaims his adherence to none of them, and claims that all good psychologists will change their point of view from that of one school to that of another if their facts and problems seem to indicate that this is desirable. It follows at once that data and problems must be collected from all possible sources, and to this principle the book remains true to the end.

All the early chapters deal with varied forms of conational activity. Since these constitute a field in which experimental research is necessarily somewhat backward, the student is, from the beginning, plunged into a consideration of theoretic interpretation. This must not be taken to mean that Mr Sprott has little regard for facts, but only that the first sorts of facts that he faces are rather general observational ones than of the experimental kind. He writes of The Springs of Action, The Social Environment, The Innate Constitution of Man, and goes on to consider Development, Conflict and Psychopathology. The regard for theoretical construction continues in his chapters on Other People and the Self, and it is only when, in Chapter XII, he arrives at a study of The Perceptual Field, and then goes on to a brief study of sensational reactions and to a longer consideration of the higher mental processes, that he shows himself thoroughly well-equipped to lead the student through a mass of experimental material.

In fact there is much to be said for this order of presentation of psychological material. Students who have not already had a reasonably good training in scientific method in other fields are much more likely to be interested and efficient if they are induced first to think psychologically about the events and problems of daily life and current discussion. It is very difficult to imagine that any student can read this book through, or that any member of the general public can dip into its pages, without being held and interested and helped to think in a genuinely psychological manner. If it has any outstanding point of difficulty it is that rather too many pieces of scattered research are mentioned with inadequate comment. Considering the condensation that is achieved it is remarkably accurate. It is throughout alive and attractive and Mr Sprott is warmly to be congratulated upon his achievement.

Explorations in Personality. By the Workers at the Harvard Psychological Clinic, H. A. MURRAY *et al.* Oxford University Press: Humphrey Milford. 1938. Pp. xiv+761. 42*s.* net.

It would be difficult to overrate the value of this study of personality which has been carried out at the Harvard Psychological Clinic. Forty-eight men of undergraduate age were studied (twenty-four intensively) by about thirty psychologists using a large number of different methods of approach—clinical, observational and experimental. The results for each subject were then collated by a 'Diagnostic council' of five, and from these the nature of his personality was deduced and described in terms of a previously established scheme.

This scheme was based upon two main procedures: (*a*) A list was drawn up of the main 'needs' or drives—achievement, dominance, aggression, etc.—together with certain other important personality variables—emotionality, various tendencies, comprising introversion and extraversion, the nature of the super-ego and the ego-ideals etc.—The subject was graded on each of these variables from observations, answers to questionnaires, etc. (*b*) A 'unity thema' and various subordinate themas were worked out for each subject, mainly from his autobiography, from clinical observation and from the phantasies which appeared in various 'imaginal productivity' tests ('thematic apperception', ink-blot, story writing, 'musical reverie', etc.). This 'unity

thema' was "the key of the individual's unique nature"..."a compound of interrelated...dominant needs" and their relation to the important stimulus situations of the subject's childhood and youth.

Though criticism may be levelled at the particular drives and themas selected, there is no doubt that they afford a very valuable and suggestive means for the combined analysis and synthesis of personality. The terminology used is sometimes to be deplored. It may also be objected that the psychoanalytic interests of four of the five members of the Diagnostic council have given an unduly Freudian bent to the themas; and indeed it would seem that the 'Case of Earnst', quoted in full, could be equally well interpreted without recourse to Freudian hypotheses. Prof. Murray, however, has in the Conclusions forestalled and to some extent met criticism on these grounds.

But it is particularly with regard to method that the authors and experimenters, especially Prof. Murray, are to be congratulated; firstly, for the collaboration of experimentalists, psychometrists and clinicians, and their attempts to dovetail their results to produce a coherent whole, instead of merely contradicting one another (this collaboration extends to the writing of the book); secondly, for the study of each individual as far as possible as a whole, rather than the investigation of isolated functions in a variety of individuals. It is to be hoped that this study will be the forerunner of others; though it may be difficult to find a similar group of psychologists who can collaborate without losing their individual outlooks and methods of approach.

The high price of the volume is to some extent necessitated by its great length and fine production; but it may unfortunately prevent many of those who would most profit from its study from buying it.

The Psychology of Social Movements: a Psycho-analytic View of Society. By Pryns Hopkins. London: George Allen and Unwin, Ltd. 1938. Pp. 284. 10*s.* 6*d.* net.

Dr Hopkins summarizes the six principal human needs as: (1) Enjoyment of knowledge, (2) Enjoyment of sensory naturalism, (3) Enjoyment of inner peace, (4) Enjoyment of the family, (5) Enjoyment of material means, and (6) Enjoyment of safety from violence. He then considers the means of fulfilment of these needs and the nature of the social and psychological forces which cause their frustration. The point of view from which these problems are considered is throughout psychoanalytical. The author considers that the habit of smoking tobacco plays a large part in the frustration of man's efforts for happiness.

The Mental Health Services of Oxford City, Oxfordshire and Berkshire. By E. F. Pinsent. Oxford: the University Press. 1937. Pp. 87. 2*s.* net.

This quite admirable report on the Oxford Mental Health services is an advance section of the Oxford Survey of Social Services (now published) carried out under the auspices of a Standing Committee of Barnett House. The report describes clearly the Statutory obligations of the Mental Health services, and an investigation of the steps taken to carry them into practice. There have been many valuable beginnings in Oxford City, in spite of lack of encouragement—not to say actual impediment—from the Central Authority. Oxfordshire and Berkshire are more backward. The following statements are of particular interest to psychologists: "We are of opinion that it needs a psychologist visiting these (retarded) classes periodically and examining the children with the class teacher in order to arrive at a valid conclusion as to the most appropriate treatment. A psychologist would also be able to give the teachers information and advice which we think they would greatly appreciate." "It is... unfortunate that all regular observation ceases when a boy or girl leaves school, for

it is just at this period that (psychological) help is often necessary to meet the troubles of adolescence." "Steps should be taken to establish a thorough and systematic periodical search in the elementary schools for mentally defective children, and then to provide a special school with suitable and sufficient accommodation for them. We think this search should be carried out by a medical psychiatrist who has special knowledge of teaching, or by an educational psychologist." "There can scarcely be two opinions as to the necessity of having one man in each (Local Government) district with special qualifications, whose opinions can be taken as authoritative, both as regards diagnosis and appropriate treatment, and whose duty it would be to co-ordinate the mental services. He should be a trained psychiatrist, and not only one of the medical staff who has picked up a desultory knowledge of the subject."

Health and Unemployment. By LEONARD C. MARSH, A. GRANT FLEMING and C. F. BLACKLER. Oxford University Press: Humphrey Milford. 1938. Pp. xxv+243. 10*s.* 6*d.* net.

This is a detailed and interesting study of the effect of unemployment upon the physical health of juveniles and adults, and the families of the latter, in Montreal. Unfortunately, it was not found possible to follow up the few suggestive hints as to the psychological outlook of the unemployed, and the effect of unemployment upon their mental health. Useful suggestions are made as to improvements in nutrition and health services.

A Visual Motor Gestalt Test and its Clinical Use. By L. BENDER. Research Monographs No. 3. New York: American Orthopsychiatric Association. 1938. Pp. ix+176.

This interesting book gives an account of the clinical use of a test of copying visually perceived figures. The figures used, eight in number, are of the line and dot variety used by Wertheimer to demonstrate *Gestalt* factors in perceptual organization, and are chosen from the examples given by him in his paper "Untersuchungen zur Lehre von der Gestalt" (1923). These are presented to the patient with the instruction to copy them. In attempting to do so, it is found, he will display forms of breakdown of visual, motor, and other functions which are characteristic of his disorder. In some cases the breakdown involves specifically *Gestalt* factors of organization; in others psychopathic tendencies, such as substitution, fractionation, and elaboration, are manifested. The forms of failure to carry out the test correctly which are found in various organic and psychogenic disorders are described, and the value of the work is much increased by the inclusion of material relating to normal children and adults, with norms for various ages. Dr Bender has opened up a line of clinical research which may come to possess considerable importance, even though the assumptions underlying it have to be modified in the course of time.

Constructing Tests and Grading. By H. D. RINSLAND. London, etc.: George G. Harrap and Co., Ltd. 1938. Pp. xvi+323. 10*s.* 6*d.* net.

This book contains a thorough and very detailed account of 'short answer' tests and examinations of every kind and in every subject, and of the points to be observed in constructing new tests. To those who use this kind of test it will be of value. The detailed discussion of methods of scoring the usual type of examination question and of grading pupils from its results should be of interest to those in charge of large public examinations. But a teacher examining a single class should approach it with caution, since he cannot assume that the performances of a small sample of children will show

a normal distribution. From the author's further dictum that "it is only through the experience and improvement of our measuring instruments of to-day, in the simpler phases of human learning, that we can push forward our scientific frontiers to a more accurate and comprehensive description of the personality of the whole child—the new centre of the educational world", we must, however, beg leave to dissent.

Zwillinge und Schule. Von HERBERT GRAEWE. Erfurt: Kurt Stenger. 1938. Pp. 132. RM. 3.60.

This monograph is a study of identical and fraternal twins. It shows how the former are not only physically alike, but also alike in personality, in interests, and in performance in various school subjects. The fraternal twins, however, show marked differences in these respects. The resemblances and differences between the children's drawings are particularly interesting.

Progressive Relaxation. By EDMUND JACOBSON. Illinois: University of Chicago Press; England: Cambridge University Press. 1938. Pp. xvii+494. 22*s.* 6*d.* net.

The first edition of this study was published in 1929 and it attracted considerable attention. The present edition describes new methods for measuring muscular contractions and general nervousness, and purports to describe also relatively exact ways in which the amounts of mental activities present in a patient at a given time may be estimated. It devotes special attention to electro-physiological measurements and technique, and maintains that by the use of these it may be conclusively demonstrated that "during progressive muscular relaxation imagery and thinking processes dwindle and disappear". The clinical significance and therapeutic use of the author's methods, both in a general and in a specific sense, are fully discussed, with some case material. A full bibliography is appended.

Although the book may, at various points, run rather beyond the evidence which is presented, it is a most serious study and deserves careful critical attention and test.

Studies in the Philosophy of Religion. By A. A. BOWMAN. Edited with a memorial introduction by NORMAN KEMP SMITH. London: Macmillan and Co. 1938. Two vols. Pp. xlviii+423; xiii+438. 30*s.* net.

Although these two volumes are, as the title indicates, addressed mainly to philosophers, they contain much that will be of interest to any thoughtful psychologist who has a bent towards critical theory. The second part, in particular, which deals rather fully with *The Anthropological Picture of Religion*, will be found to possess psychological significance, and also the discussions dealing with so-called naturalistic ethics. On the other hand the late Prof. Bowman does not display any detailed knowledge or interest in recent psychological contributions to the study of religious phenomena.

Über Sinn und Sinnlosigkeit des Lebens. Von MARGARETHE v. ANDICS. Wien: Gerold und Co. 1938. Pp. iii+175.

In Vienna, those who are rescued from attempted suicide are (or were) taken to the University Clinic for Psychiatry and Neurology, for observation and treatment. A hundred of these cases were questioned by the author about their life histories and their reasons for suicide, with a view to discovering what it is that makes men feel that their lives are no longer worth living; how their existence becomes meaningless to them, and they seem to be of no use to anyone or for anything. Though the method adopted is psychological, the author's bent and her conclusions are mainly philosophical in nature.

Dasein und Wirklichkeit: Eine Einführung in die Seinslehre. Von Otto Janssen. München: Ernst Reinhardt. 1938. Pp. 351. RM. 8, Leinen RM. 10.

Seele als Äusserung: Untersuchungen zum Leib-Seele-Problem. Von Paul Helwig. Leipzig und Berlin: G. B. Teubner. 1936. Pp. 124. RM. 4.20; f. Ausland RM. 3.15.

Neue Dialoge zwischen Hylas und Philonous. Von H. A. Wimmer. Heidelberg: Carl Winter's Universitätsbuchhandlung. 1938. Pp. viii+154. Kart. RM. 7.

Drei Vorträge über Philosophie und Parapsychologie. Von Johannes J. Poortman. Leiden: A. W. Sitjhoff's Uitgeversmij N.V. 1939. Pp. 76.